BEITRÄGE ZU THEORIE UND ANWENDUNG VON PERMUTATIONSTESTS IN DER UNI- UND MULTIVARIATEN DATENANALYSE

Klaus Willmes

Disbui. Trier
1987

UNIV.
LEIDEN
NL

BEITRÄGE ZU THEORIE UND ANWENDUNG VON PERMUTATIONSTESTS IN DER UNI- UND MULTIVARIATEN DATENANALYSE

Dem Fachbereich I der Universität Trier

vorgelegte Dissertation zur

Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von Klaus Willmes

aus

Neheim-Hüsten (Arnsberg 1) / NRW

Gutachter:

Professor Dr. J. Bredenkamp,

Professor Dr. D. Bartussek,

Professor Dr. E. Sonnemann

Tag der mündlichen Prüfung: 17. Juli 1987

Hiermit möchte ich mich bei Herrn Professor Dr. J. Bredenkamp, Herrn Professor Dr. D. Bartussek und Herrn Professor Dr. E. Sonnemann sehr herzlich für die Bereitschaft bedanken, diese Arbeit zu betreuen.

Ohne die Dissertation meines ehemaligen Kollegen Dr. rer. nat. Norbert Pyhel an der Abteilung für Medizinische Statistik und Dokumentation der RWTH Aachen und unsere gemeinsame Arbeit über univariate Permutationstests wäre diese Dissertation nicht möglich gewesen. Auch die ehemaligen Kollegen und Kollegin Frau Dipl. Math. Gudrun Henn (Aachen), Dr. Guido Giani (Aachen) und Privatdozent Dr. rer. nat. Peter Roebruck (Heidelberg) haben durch viele hilfreiche Diskussionen und eigene Arbeiten über Permutationstests wesentlich zur endgültigen Form dieser Arbeit beigetragen. Alle verbliebenen Fehler und Unklarheiten gehen zu meinen Lasten. Professor Dr. Rudolf Repges, Vorstand der Abteilung Medizinische Statistik und Dokumentation an der RWTH Aachen, hat die Arbeit stets unterstützt und mir die Rechnerkapazität seines Institutes ohne Einschränkung zur Verfügung gestellt.

Ohne die großen Programmierkenntnisse meines Mitarbeiters im Projekt der DFG - Forschergruppe "Aphasie und kognitive Störungen", des mathematisch-technischen Assistenten Herbert Ratajczak wären die verschiedenen Permutationstestalgorithmen kaum als fertige Programme in diese handhabbare Form gekommen.

Professor Dr. Klaus Poeck, Vorstand der Abteilung Neurologie an der RWTH Aachen hat mir die großzügige Gelegenheit geboten, über mehrere Jahre in der interdisziplinär zusammengesetzten neurolinguistischen und neuropsychologischen Arbeitsgruppe dieses methodische Thema bearbeiten zu können. Meine Kolleginnen und Kollegen in dieser Arbeitsgruppe Dr. Ria De Bleser, Professor Dr. Wolfgang Hartje, Professor Dr. Walter Huber, Dr. Walter Sturm und Dr. Dorothea Weniger haben mir durch ihre eigenen und gemeinsame experimentelle Untersuchungen eine Fülle von Daten geliefert, an denen ich die Permutationstests erproben konnte.

Schließlich gilt mein Dank Frau Elisabeth Böhler, Frau Elisabeth Gaeta und Frau Kathi Seidel in der Abteilung Neurologie, die verschiedene Fassungen des Manuskripts mit großer Ausdauer geschrieben haben.

The logical reason for randomization is that it is possible on the basis of a model reflecting the randomization to draw sound statistical inferences the probability basis of the model being provided not by wishful thinking but by the actual process of randomization which is part of the experiment.
(H. Scheffe, 1959)

The simple fact that the act of random treatment allocation in itself allows valid statistical inference does not appear to have penetrated the collective consciousness of statisticians. It is trivially true that the validity of randomization tests is general and does not require the observations to be independent or, indeed, to be random variables at all.
(K.R. Gabriel & W. J. Hall, 1983)

Permutation tests form a large vertebra in the backbone of nonparametric statistics.
(R.G. Miller, 1981)

INHALTSVERZEICHNIS Seite

1.	EINFÜHRUNG UND ÜBERBLICK	1
2.	DAS PRINZIP DES PERMUTATIONSTESTS UND DES RANDOMISIERUNGSTESTS	26
3.	PERMUTATIONS- UND RANDOMISIERUNGS-TESTS FÜR LINEARE HYPOTHESEN IN LINEAREN MODELLEN - UNIVARIATER FALL	36
3.1.	Invarianz-Permutationstests	37
3.1.1.	Reduktion eines Testproblems durch Invarianz	37
3.1.2.	Beispiel für die Reduktion durch Invarianz	40
3.1.3.	Ableitung eines Permutationstests für lineare Hypothesen im linearen Modell	42
3.1.4.	Ein Invarianz-Permutationstest für den einfaktoriellen Blockplan	44
3.1.5.	Invarianz-Permutationstests für verschiedene ein- und mehrfaktorielle Versuchspläne	47
3.1.6.	Zusammenfassung und Diskussion	95
3.2.	Diskussion des Konstruktionsprinzips für Invarianz-Permutationstests	97

		Seite
3.3.	Randomisierungstests für lineare Hypothesen in linearen Modellen	103
3.3.1.	Populationsmodell oder experimentelle Randomisierung	103
3.3.2.	Verschiedene Darstellungen des Randomisierungsmodells	105
3.3.3.	Tests linearer Hypothesen in Randomisierungsmodellen	110
3.3.4.	Existenz von Invarianz-Permutationstests in Randomisierungsmodellen	115
4.	INVARIANZ-PERMUTATIONSTESTS FÜR LINEARE HYPOTHESEN IN LINEAREN MODELLEN - MULTIVARIATER FALL	124
4.1.	Multivariate Experimente	124
4.2.	Ansätze zu Permutationstests für multivariate Experimente	125
4.3.	Übertragung des Konstruktionsprinzips für Invarianz-Permutationstests bezüglich linearer Hypothesen in linearen Modellen auf den multivariaten Fall	132
4.4.	Beispiel: Test auf Wechselwirkungen im multivariaten Split-Plot Plan	138
4.5.	Multivariate Invarianz-Permutationstests für ein- bis dreifaktorielle Versuchspläne	145
4.6.	Multivariate Permutationstests zur Analyse von Profilen	157

		Seite
4.7.	Multivariate Permutations/Randomisierungstests mit normierten Distanzen zwischen Beobachtungen als Analyseeinheiten (MRPP-Statistiken)	169
4.7.1.	(Un)verbundenes multivariates k-Stichprobenproblem	169
4.7.2.	Verallgemeinerung des MRPP-Ansatzes auf mehrfaktorielle multivariate Versuchspläne	175
5.	MULTIPLE PERMUTATIONSTESTS	178
5.1.	Bekannte Verfahren für uni- und multivariate multiple Permutationstests	179
5.1.1.	Verfahren, die auf Zweistichproben-Permutationstests basieren	179
5.1.2.	Verfahren, die auf den bei Gültigkeit der globalen Nullhypothese zulässigen Permutationen basieren	181
5.1.3.	Gabriel's simultane Testprozedur (STP) für simultane Permutationstests	186
5.1.4.	Inadäquatheit gewisser simultaner Testprozeduren für Permutationstests	188
5.2.	Multiple Tests - Definitionen und theoretische Ansätze	190
5.3.	Multiple Permutationstests für die Komponenten multivariater Beobachtungen	203

		Seite
5.3.1.	Multiple sequentielle Permutationstests zum multiplen Niveau alpha	203
5.3.2.	Simultane Testprozeduren für Permutationstests	218
5.4.	Multiple sequentielle Permutationstests für die Faktorstufen in uni- und multivariaten Versuchsplänen zum multiplen Niveau alpha	221
5.4.1.	Vergleiche aller Faktorstufen	224
5.4.2.	Vergleiche gegen eine Kontrolle	233
5.5.	Simultane Permutationstests für die Faktorstufen in uni- und multivariaten Versuchsplänen zum globalen Niveau alpha	236
5.6.	Multiple Permutationstests für Hypothesenfamilien bezüglich Komponenten multivariater Beobachtungen und bezüglich Faktorstufen	239
6.	TESTSTÄRKE UND ROBUSTHEIT VON PERMUTATIONS- UND RANDOMISIERUNGSTESTS	245
6.1.	Probleme bei der Bestimmung der Teststärke	245
6.1.1.	"Optimale" Permutationstests	245
6.1.2.	Unmöglichkeit einer apriori Festlegung der Teststärke	246
6.1.3.	Unmöglichkeit der Festlegung einer durchschnittlichen Effektstärke	248

		Seite
6.1.4.	Teststärke von multiplen Testprozeduren	250
6.2.	Bekannte Verfahren und Ergebnisse für Permutationstests	252
6.2.1.	Asymptotische Teststärke	252
6.2.2.	Approximation der finiten Teststärke	254
6.2.3.	Finite konditionale und unkonditionale Teststärkebestimmung	255
6.2.4.	Inversion von Permutationstests	260
6.3.	(Approximative) Bestimmung der finiten konditionalen und unkonditionalen Teststärke	270
6.3.1.	Exakte und geschätzte konditionale Teststärke für Permutations- und Randomisierungstests	270
6.3.2.	"Naive" Bestimmung der finiten konditionalen Teststärke	275
6.3.3.	Approximative apriori Bestimmung der finiten unkonditionalen Teststärke von Permutationstests	277
6.3.4.	(Approximative) Bestimmung der Teststärke von multiplen Testprozeduren	280
6.3.5.	Diskussion	284
6.4.	Robustheit von uni- und multivariaten Randomisierungs- und Permutationstests	286
6.4.1.	Definitionen von Robustheit	286
6.4.2.	Univariate Randomisierungs- und Permutationstests	288

		Seite
6.4.3.	Multivariate Randomisierungs- und Permutationstests	292
6.4.4.	Diskussion	294
7.	BEDEUTUNG VON RANDOMISIERUNGS- UND PERMUTATIONSTESTS FÜR DIE UNI- UND MULTIVARIATE DATENANALYSE IN DER PSYCHOLOGIE	296
7.1.	Bedeutung für eine deduktivistische Theorie des Experimentierens	296
7.2.	Bedeutung für die Legitimierung statistischer Hypothesentests	304
7.3.	Das Testen der Effekte aller unabhängigen Variablen in uni- und multivariaten Versuchsplänen	312
7.3.1.	Problematische Aspekte dieses Vorgehens	312
7.3.2.	Univariate mehrfaktorielle Experimente	314
7.3.3.	Multivariate mehrfaktorielle Experimente	322
7.4.	Verfahren zur möglichst strengen Prüfung wissenschaftlicher Hypothesen mit Signifikanztests	325
7.4.1.	Überblick	325
7.4.2.	Einige typische Fragestellungen	332

		Seite
7.4.3.	Permutationstests für (streng) monotone Trend-Hypothesen bezüglich qualitativ abgestufter unabhängiger Variablen	355
7.4.4.	Tests zur Analyse des Typs einer Wechselwirkung im linearen Modell	372
8.	PRAKTISCHE AUSFÜHRUNG VON RANDOMISIERUNGS- UND PERMUTATIONSTESTS	388
8.1.	Überblick über bekannte Verfahren	388
8.2.	Die Monte-Carlo Version des Permutationstests	396
8.3.	Allgemeiner Aufbau von (Invarianz-) Permutationstest-Programmen	404
8.4.	Spezielle Programme für Invarianz-Permutationstests	429
8.4.1.	Univariate Permutationstests	430
8.4.2.	Multivariate Permutationstests	438
8.4.3.	Bestimmung der Teststärke	443
8.4.4.	Multiple Tests für die Komponenten multivariater Beobachtungen	449
8.4.5.	Multiple Tests für Faktorstufen	457
	LITERATURVERZEICHNIS	469

Seite

ANHANG 504

Überblick 504

A1 Bezeichnungen 505

A2 Eigenschaften verschiedener Matrizenoperationen 510

A3 Eigenschaften von Permutationsgruppen 513

A4 Lineare Unterräume und Projektionen 514

B1 FORTRAN-Version des Shift-Algorithmus von Streitberg & Röhmel für das unverbundene Zweistichprobenproblem 515

B2 FORTRAN-Version des Shift-Algorithmus von Streitberg & Röhmel für das Einstichproben- bzw. verbundene Zweistichprobenproblem 524

B3 FORTRAN-Listing der sequentiellen multiplen Testprozedur für die abhängigen Variablen in einem multivariaten Testproblem: Subroutine SEQMTP 531

B4 FORTRAN-Listing des Programmes für den Test auf Wechselwirkungen im multivariaten Split-Plot Versuchsplan 535

B5 Verzeichnis der vorhandenen FORTRAN-Programme 557

Das Prinzip des Randomisierungs- bzw. Permutationstests ist von R.A. Fisher (1935) in seinem Buch "The Design of Experiments" zum ersten Mal ausführlich behandelt worden. Permutationstests sind geeignet, auch dann valide Signifikanztests zu ermöglichen, wenn keine Zufallsstichproben aus definierten Populationen vorliegen, sondern wenn mittels zufälliger Zuteilung (Randomisierung, Randomisation) Beobachtungseinheiten verschiedenen experimentellen Bedingungen zugewiesen werden. Dieses Prinzip der experimentellen Randomisierung hatte R.A. Fisher in dem Aufsatz "The arrangement of field experiments" (1926) betont. Nach Scheffe (1959, Fußnote S. 291 und auch Picard, 1978) hat Neyman schon 1923 ein Randomisierungsmodell für den vollstandigen Zufallsplan (Neyman, 1923) und 1935 für den zufälligen Blockplan und das Lateinische Quadrat (Neyman, 1935) formuliert. Obwohl der Permutationstest die Methode "of examining the wider hypothesis in which no normality of distribution is implied" (Fisher, 1935, S. 44) war, betrachtete Fisher wegen des i.a. notwendigen Rechenaufwandes dieses Verfahren als unpraktisch. In seinen "Statistical Methods for Research Workers" (1925 und spätere Auflagen) wird es auch nicht erwähnt.

Pitman (1937 a, b; 1938) und Welch (1937) umgingen dann auch die Ermittelung der exakten Permutationsverteilung ihrer Teststatistiken und bestimmten lediglich die ersten vier Momente der exakten Permutationsverteilung. Sie konnten zeigen, daß diese mit wachsendem Stichprobenumfang gegen die entsprechenden Momente bekannter Verteilungen

(Normalverteilung, Chi-Quadrat-Verteilung, Beta-Verteilung) konvergierten.

Eine mathematisch befriedigende Formulierung der Permutationstestprinzipien gelang Scheffe (1943a, b). Er wies nach, daß Permutationstests - qua Konstruktion - streng verteilungsfreie ähnliche Tests zum Niveau alpha für bestimmte Permutationsinvarianz - Nullhypothesen sind. Die Arbeiten von Lehmann und Stein (1949) und Hoeffding (1952) entwickelten die Theorie weiter (siehe auch Lehmann, 1959, Kap. 5.7-5.10 und Pratt und Gibbons, 1981, Kap. 4,6). Wald & Wolfowitz (1944) waren die ersten, die einen Satz über die Konvergenz der Permutationsverteilung von linearen Permutationsteststatistiken beweisen konnten (vgl. auch Silvey, 1953, 1954).

Kempthorne (1952, 1955, 1966a) und seine Schüler Wilk (1955) und Zyskind (1962, 1963) sowie Wilk & Kempthorne (1955, 1956) analysierten für viele verschiedene ein-und mehrfaktorielle Versuchspläne die Möglichkeiten einer Theorie für das lineare Modell, in der nur die experimentelle Randomisierung als stochastisches Element in die Modellgleichungen eingeht (sog. finites Modell). Unter diesen Bedingungen untersuchten sie die Eigenschaften von Schätzwerten für feste und zufällige Behandlungseffekte, ihre Erwartungswerte, Varianzen und Kovarianzen. Ebenso konnten die Erwartungswerte der im parametrischen linearen Modell (üblichen) mittleren Quadratsummen hergeleitet und Nullhypothesen bezüglich bestimmter Behandlungseffekte getestet werden. Dazu wurden jedoch (wiederum) nicht die exakten Permutationsverteilungen bestimmt, sondern die F-Verteilungen der Teststatistiken aus dem gängigen linearen

Modell mit Normalverteilungsannahme als hinreichend gute Approximationen an die Permutationsverteilungen angesehen (z.B. Kempthorne, 1955, S. 966). Bekannte Studien zur Robustheit für große Stichprobenumfange der F-Statistiken im Randomisierungsmodell sowohl unter der jeweiligen Nullhypothese wie unter Verschiebungsalternativen sind die von Box & Andersen (1955) und Baker & Collier (1966b) sowie Collier & Baker (1966). Für den praktisch so wichtigen Fall kleiner Stichprobenumfänge ist wenig bekannt. Verschiedentlich wird vermutet, daß es hier zu großen Abweichungen kommen kann.

Diese Arbeiten von Kempthorne und Mitarbeitern sind in den gängigen Lehrbüchern über die varianzanalytische Auswertung von Versuchsplänen kaum oder gar nicht berücksichtigt worden (Kirk, 1968; Winer, 1971). Falls sie behandelt wurden, dienten sie lediglich als weitere Rechtfertigung für das übliche varianzanalytische Modell. Eine Ausnahme bildet das Lehrbuch von Scheffe (1959, Kap. 9), das einen Überblick über das Permutationstest-Konzept, Verfahren für gängige Versuchspläne und offene Probleme gibt. Lehrbücher zur nichtparametrischen Statistik behandeln Permutationstests häufiger (s. z.B. Siegel, 1956; Bradley, 1968; Conover, 1971; Lienert, 1973).

Weitere theoretische Arbeiten in den 50er und 60er Jahren vertieften das Verständnis über Eigenschaften von Permutationstests (z.B. Lehmann, 1959). Optimalitätseigenschaften von Permutationstests konnten von Lehmann allerdings nur für den - praktisch wenig interessanten - Fall gezeigt werden, in dem die betrachtete Alternative eine Normalverteilungsalternative oder eine

andere vollstandig spezifizierte Alternative ist. Das ist bis heute die einzige formale Methode geblieben, einen optimalen Permutationstest auszuwahlen, der in diesem Fall mit der üblichen Teststatistik des parametrischen linearen Modells ausgeführt wird (s. Witting, 1969; Witting & Nolle, 1970; Gabriel & Hall, 1983). Einen sehr knappen überblick uber den Stand der theoretischen Ergebnisse bis 1970 geben Puri & Sen (1971, Kap. 3.3-3.5) zu Permutationsgrenzwertsätzen und asymptotischer Schärfe von Permutationstests.

Dwass (1957) und unabhängig davon Barnard (1963) und Hope (1968) schlagen vor, nicht alle Permutationen zur Bestimmung der exakten Permutationsverteilung der jeweiligen Teststatistik zu nutzen, sondern mit einer (großen) Zufallsstichprobe (mit oder ohne Zurücklegen) von Permutationen (sog. Monte-Carlo Methode) diese zu "approximieren" (sog. approximative Permutationstests). Dieser Ansatz machte den Weg frei, Permutationstests auch tatsächlich zu benutzen. Die Bezeichnung approximative Permutationstests ist allerdings irreführend, denn auch der Test mit einer Zufallsauswahl ist ein exakter Test zum Niveau alpha, wie Dwass gezeigt hat (s. auch Pagano & Tritchler, 1983). Außerdem konnte gezeigt werden, daß die Einbuße an Testschärfe nicht gravierend ist (Vadiveloo, 1983). Den heutigen Forschungsstand zu diesem Thema geben die Arbeiten von Jöckel (1982, 1986a, b) wieder.

Ohne die Fortschritte in der Rechengeschwindigkeit der Computer wäre aber selbst die Durchführung der Monte-Carlo Version des Permutationstests (meist mit 1000 - 20000 zufällig ausgewählten Permutationen) nicht praktikabel. Erst ab Ende der 70er Jahre werden Computer-Programme für verschiedene Versuchspläne publiziert (Green, 1977; Hawel, 1982; Heller, 1981a, b; Pagano & Tritchler, 1983; Soms, 1977; Still & White, 1981; Streitberg & Rohmel, 1983, 1984, 1986).

Gleichzeitig wächst die Zahl von Publikationen, in denen Permutationstests für immer neue Versuchspläne und Fragestellungen vorgeschlagen werden.

Mit dem Problem von Mehrfachvergleichen oder simultanen Testprozeduren haben sich nur wenige Autoren beschäftigt. Miller (1981), Shuster & Boyett (1979) und auch Gabriel (1979) schlagen Verfahren vor, die in direkter Übertragung "parametrische" Teststatistiken verwenden. Petrondas & Gabriel (1983) konnten jedoch zeigen, daß die Fehlerwahrscheinlichkeit für den Fehler 1. Art größer als das nominelle Signifikanzniveau ausfallen kann. Da Hypothesen über die Gleichheit einer Teilmenge von Versuchsbedingungen nur die Permutationsinvarianz uber die eingeschränkte Menge von Permutationen innerhalb dieser Teilmenge von Bedingungen implizieren, dürfen diese Teststatistikwerte nicht mit der Verteilung der Teststatistik über alle Bedingungen verglichen werden. Eine Lösungsmöglichkeit stellen (sequentielle) multiple Tests wie bei Marcus, Peritz & Gabriel (1976), Begun & Gabriel (1981) und Sonnemann (1981a, b; 1982) dar. Auf sie wird im Abschnitt 5 eingegangen.

Auch die Anwendungen in verschiedenen Wissenschaftszweigen nehmen zu, in denen parametrische Modelle mit Normalverteilungsannahmen umstritten oder unhaltbar sind; exemplarisch seien Untersuchungen zur kunstlichen Stimulierung von Regenfällen (Gabriel, 1979; sowie das ganze Heft 8 der Communications in Statistics, 1979) und klinische Studien zum Vergleich verschiedener Behandlungsmethoden (Matts & McHugh, 1978, 1983 ; McHugh & Matts, 1983; Shen & Quade, 1983) genannt.

Neuere Studien zur asymptotischen Teststärke stammen von Robinson (1973a) und verbesserte Approximationen an die Verteilung unter der Nullhypothese für das Ein- und Zweistichprobenproblem von Robinson (1978, 1983) und für das Mehrstichprobenproblem von Robinson (1980).

Mitte der 70er Jahre kommt es, vielleicht auch angeregt durch die größere Zahl an Publikationen, zu einer Diskussion über den Stellenwert der experimentellen Randomisierung (Kontra: Harville (1975), Basu (1980); Pro: Kempthorne (1975, 1977)) und über das Verhältnis von Signifikanztests und Tests von Hypothesen (Kempthorne, 1976, 1983 und ein ganzes Heft der Communications in Statistics, 1976). Der zweite Problemkreis ist deshalb von Bedeutung, weil wenig über die Teststärke von Permutationstests gegenüber verschiedenen Arten von Alternativen bekannt ist und weil Permutationstests bedingte Tests sind, sodaß eine exakte Berechnung der Teststärke vor Durchführung eines Experimentes streng genommen nicht moglich ist. Methoden zur Bestimmung der bedingten Teststärke sind z.B. bei Kempthorne & Doerfler (1969), Gabriel (1978), Gabriel & Hall (1983), Gabriel & Hsu (1983) zu finden. Auf dieses Problem wird

später in Abschnitt 6 erneut eingegangen.

Uneinheitlich ist auch die Bezeichnung für das von Fisher begründete Testverfahren. Kempthorne & Doerfler (1969) schlagen vor, den Namen "Permutationstest" für den Fall zu reservieren, in dem Zufallsstichproben aus nicht bekannten (oder unspezifizierten) Verteilungen miteinander verglichen werden und von einem "Randomisierungstest" nur dann zu sprechen, wenn Ergebnisse von Versuchseinheiten verglichen werden, die aufgrund experimenteller Randomisierung verschiedenen Bedingungen zugewiesen wurden. Auf diese Unterscheidung wird im Abschnitt 3.3. noch eingegangen. Der Name "re-randomisation test" (Gabriel & Hall, 1983) verdeutlicht noch stärker den Aspekt der experimentellen Randomisierung: Zuerst wird ein bestimmter Versuchsplan aus einer Menge von möglichen Versuchsplänen zufällig ausgewählt und die Versuchseinheiten den verschiedenen experimentellen Bedingungen zufällig zugeteilt. Um die interessierende (Null-) Hypothese zu testen, werden alle bei Gültigkeit der Nullhypothese möglichen (zulässigen) Zuteilungen der Beobachtungen zu den Bedingungen vorgenommen, um den ursprünglichen Wert der gewählten Teststatistik mit ihren Werten aufgrund der Re-Randomisierung zu vergleichen. Diese Re-Randomisierungen sind aber nichts anderes als Permutationen der Komponenten des Vektors der Beobachtungen. Deshalb spricht Robinson (1973a) auch von "permutation tests for randomisation models".

Bisher ist ausschließlich über Permutationstests berichtet worden, bei denen je Versuchs-/Beobachtungseinheit und experimenteller Bedingung nur eine (abhangige) Variable (univariates Experiment) erhoben worden ist. Eine Verallgemeinerung auf mehrere gleichzeitig zu berücksichtigende Variablen je Versuchseinheit (d-dimensionale Beobachtungen, multivariates Experiment) ist aber ohne konzeptuelle und theoretische Probleme moglich (Vgl. Puri & Sen, 1971, S. 68-69). In diesem Fall werden die Versuchseinheiten zum Testen der jeweiligen Hypothese mit ihren d-dimensionalen Beobachtungswerten den verschiedenen Bedingungen zufällig zugewiesen. Diese Erweiterung haben als erste Wald & Wolfowitz (1944) für das unverbundene Zweistichprobenproblem für d=2 Variablen vorgenommen. Sie weisen nach, daß eine monotone Transformation der im parametrischen Fall üblichen T^2-Statistik von Hotelling (1931) unter schwachen Bedingungen asymptotisch eine Chi-Quadrat-Verteilung mit d=2 Freiheitsgraden besitzt. Chung & Fraser (1958) untersuchen dasselbe Testproblem für beliebiges d, benutzen aber Teststatistiken, die aus der Summe d univariater Teststatistiken gebildet sind und so nicht die mögliche Kovariation der abhängigen Variablen explizit berücksichtigen. Arnold (1964) behandelt das verbundene Zweistichprobenproblem in d Dimensionen. Er ermittelt die exakten ersten vier Momente der Permutationsverteilung seiner Teststatistik. Royen (1978) hat gute Teststärkeeigenschaften für d=2 und Normalverteilungsalternativen berichtet. Boyett & Shuster (1977) benutzen für beide Zweistichprobenprobleme jeweils als multivariate Statistik das Maximum der univariaten Statistiken. Ihr Interesse war es, einen einseitigen Test

für den Fall zu entwickeln, daß eine Behandlung, verglichen mit einer Kontrollbedingung, für mindestens eine der abhängigen Variablen einen Effekt zeigt. Zur Testentscheidung benutzen sie ein Programm, das für kleine Stichprobenumfänge den exakten Test und für größere Anzahlen von Versuchseinheiten einen approximativen Permutationstest verwendet (Vgl. Shuster & Boyett, 1979, S. 382).

Mardia (1971) analysiert die Robustheit von Teststatistiken für das multivariate Regressionsproblem bei Verletzung der Normalverteiltheit. Dabei stellt er fest, daß es nur für das Spurkriterium $V^{(s)}$ von Pillai (1955) möglich sei, auch dessen Permutationsverteilung zu analysieren. Die Konstruktion eines exakten Permutationstests erwähnt Mardia nur kurz. Er bestimmt vielmehr die beiden ersten Momente der Verteilung von $V^{(s)}$ unter dem Permutationsmodell und unter der Annahme einer in ihren Argumenten symmetrischen gemeinsamen Verteilungsfunktion der Fehler. Das Spurkriterium ist im Spezialfall des k Stichprobenproblems mit d-variaten Beobachtungen robust, falls die k Stichprobenumfänge nur wenig verschieden sind. Mardias Ergebnisse verallgemeinern die Arbeiten von Box & Watson (1962) für das univariate Regressionsproblem, von Barton & David (1962) für den bivariaten Fall des k Stichprobenproblems und von Mardia (1967) für das bivariate Zweistichprobenproblem.

Ein weiterer Versuch, Permutationstests für multivariate Fragestellungen zu entwickeln, geht auf Mielke, Berry & Johnson (1976) zurück. Ihre MRPP-Statistiken (Multi-Response-Permutation-Procedures) sind Linearkombinationen der gemittelten (euklidischen) Abstände zwischen den d-

dimensionalen Beobachtungen innerhalb der k experimentellen Bedingungen oder Gruppen von Versuchseinheiten. Auch für den zufälligen Blockplan wurde eine MRPP-Statistik (Mielke & Iyer, 1982) publiziert und ein Computer-Programm von Iyer, Berry & Mielke (1983) vorgestellt. Zur Bestimmung der Permutationsverteilung werden die ersten drei Momente der Teststatistik bestimmt und behauptet, daß auch schon für kleine Stichprobenumfänge die exakte Permutationsverteilung zumindest an ihrem unteren Ende - denn kleine Werte der Teststatistik sprechen gegen die Nullhypothese - gut durch eine die Schiefe der Verteilung berücksichtigende Pearson-III-Verteilung angenähert werden kann (Berry & Mielke, 1983a). Auch theoretisch konnte Mielke (1979b) zeigen, daß asymptotisch unter bestimmten Bedingungen eine schiefe Verteilung der Teststatistiken resultiert. In neueren Arbeiten werden sogar die ersten vier Momente berücksichtigt (Tracy & Tajuddin, 1985).

Insgesamt ist jedoch dem Urteil von Kempthorne (1977) zuzustimmen, daß Randomsierungstests für multivariate Beobachtungen noch "in den Kinderschuhen stecken". Er vermutet auch, daß die parametrischen MANOVA-Teststatistiken wegen ihrer Nichtlinearität nicht so robust sind gegen Verletzungen der Verteilungsannahmen des parametrischen linearen Modells (Vgl. Edgington, 1980, S. 181).

In der psychologischen Methoden-Literatur sind Permutationstests wenig beachtet worden. Ein früher Übersichtsaufsatz, der sich allgemeiner mit nichtparametrischen Methoden beschäftigt, stammt von Moses (1952). Die Lehrbücher von Guilford (1956), Hays (1963),

Winer (1971) oder Bortz (1977) etwa behandeln bis auf wenige Anmerkungen nur das parametrische lineare Modell. Eine Ausnahme bilden Marascuilo & McSweeney (1977) und besonders Edgington, der, beginnend mit einer Arbeit im Journal of Psychology (Edgington, 1964), in vielen Publikationen (Edgington, 1965, 1966, 1967, 1969a, b, 1970, 1971, 1973, 1975a, b; Edgington & Ezinga, 1978; Edgington & Strain, 1973) die Bedeutung von Randomisierungstests für die psychologische Forschung betont hat. Seine Ideen und Methoden sind in dem ersten Lehrbuch, das ausschließlich Randomisierungstests behandelt und auch diesen Titel trägt, zusammengefaßt (Edgington, 1980). Kritische Besprechungen des Buches findet man von Green (1981), Kempthorne (1982) und Singer (1981). Edgington betont die Validität des Randomisierungstests, auch wenn keine Zufallsstichproben betrachtet werden. Vielmehr bezeichnet er die Annahme von Zufallsstichproben für ein Experiment als sehr selten gerechtfertigt und damit als normalerweise irrelevant (1980, S. iii); außerdem ist die tatsächliche Herstellung einer Zufallsstichprobe über eine Tabelle von Zufallszahlen oder einen Zufalls(zahlen)generator i.a. nicht praktikabel und wird von Experimentatoren selten durchgeführt. Wichtig ist allein die randomisierte Zuweisung der Beobachtungseinheiten zu den experimentellen Bedingungen. Die Generalisierung der im einzelnen Experiment gewonnenen Ergebnisse sei kein statistisches Verfahren, sondern ein induktiver Schluß. Weiterhin weist Edgington nachdrücklich auf die Schwierigkeit hin, in der Psychologie sinnvoll eine Population zu definieren, da psychologische Gesetzesaussagen sich im allgemeinen nicht auf endliche, im Augenblick des

Experimentes existierende Populationen beziehen. Bedeutung hat das Buch von Edgington auch deshalb gewonnen, weil der Autor ausführlich auf verschiedene Randomisierungstests für Einzelfallexperimente eingeht (1980, Kap. 10) und verschiedene Pläne zur randomisierten Zuweisung von verschiedenen Behandlungen vorstellt. Die Anwendung von Randomisierungstests für diesen besonders in der klinischen (Therapie-)Forschung wichtigen Fall behandeln u.a. Levin, Marascuilo & Hubert (1978), Petermann & Hehl (1979) und Möbus & Nagl (1983).

Zu problematisieren ist jedoch die Empfehlung von Edgington, zu jedem parametrischen statistischen Test für komplexe faktorielle Versuchspläne durch (zulässige) Permutationen der Beobachtungen - bei gleicher Wahl der Teststatistik - einen vollständig verteilungsfreien Test zu erzeugen (Edgington, 1980, S. 16), zumal er selber später feststellt (1980, S. 162), daß es für den zweifaktoriellen Zufallsplan keinen validen Randomisierungstest auf Wechselwirkung zwischen beiden experimentellen Faktoren gibt. Dieses Problem der Existenz von Permutationstests für lineare Hypothesen in mehrfaktoriellen Modellen wird wesentlicher Bestandteil der Arbeit sein.

Tests für multivariate Beobachtungen werden von Edgington nur sehr knapp behandelt. Neben dem Vorschlag, univariate Statistiken für Linearkombinationen der abhängigen Variablen bzw. die Summe der univariaten Statistiken zu benutzen, gibt es nur den Hinweis, daß multivariate Teststatistiken, die die Kovariation der abhängigen Variablen berücksichtigen, angemessener seien.

Pyhel (1978a, b, 1980a, b) hat einen allgemeinen Ansatz zur Konstruktion von Permutationstests in mehrfaktoriellen linearen Modellen vorgestellt. In einem mehrfaktoriellen Versuchsplan stellt sich als wichtigstes Problem, die Testentscheidung bezüglich der Effekte eines Faktors (oder dessen Wechselwirkung(en)) unabhängig von möglichen anderen vorhandenen Effekten (oder deren Wechselwirkung(en)) treffen zu können.

Pyhel benutzt das Prinzip der Reduktion eines Testproblems durch Invarianz (Lehmann, 1959; Witting & Nölle, 1970). Dieses Prinzip sichert, daß man auch dann ein zum ursprünglichen Testproblem äquivalentes Problem testet, wenn man die nicht interessierenden, "störenden" Parameter (-Schätzungen) des linearen Modells von den vorliegenden Beobachtungen abzieht (sog. Alignment) und mit einer geeigneten Teststatistik den Permutationstest durchführt. Für die Fehlervariablen des Modells werden von Pyhel schwächere Annahmen als im Normalverteilungsmodell gemacht; sie beziehen sich lediglich darauf, welche Fehlerterme im linearen Modell für den jeweiligen Versuchsplan vertauscht werden durfen. Diese "erlaubten" Vertauschungen entsprechen oft den randomisierten Zuteilungen der Versuchseinheiten auf die experimentellen Bedingungen im jeweiligen Versuchsplan. Der Ansatz von Pyhel hat also eine gewisse "Zwitterstellung" zwischen dem finiten Randomisierungsmodell bei Kempthorne und Mitarbeitern (1952, 1955, 1977) und dem üblichen Populationsmodell. Zwar macht Pyhel eine schwache Vertauschbarkeits-Verteilungsannahme und befindet sich damit konzeptuell "automatisch" in einem Populationsmodell (vgl. Lehmann, 1975, S. 56), aber die Annahmen haben große Ähnlichkeit mit den Eigenschaften der sog. Designvariablen

im Modell von Kempthorne. Thornett (1982a, b) thematisiert diesen Unterschied, in dem er zwei Arten von Statistikern gegenüberstellt: den "modeller" und den "randomizer". Thornett gelangt für die von ihm behandelten Versuchspläne zu den selben Vertauschungsannahmen wie Pyhel. Auf diesen Fragenkomplex wird im folgenden ausführlicher eingegangen (vgl. Abschnitt 3.3).

Mit den Invarianz-Permutationstests von Pyhel konnte z.B. ein verteilungsfreier, mit obigen Invarianzeigenschaften ausgestatteter Test auf Wechselwirkungen im Split-Plot Versuchsplan hergeleitet werden (vgl. Willmes & Pyhel, 1981; Willmes, 1982). Es ist aber auch möglich, daß ein Invarianz-Permutationstest für ein bestimmtes Testproblem nicht existiert. Pyhel (1978a) hat skizziert, daß ein Invarianz-Permutationstest auf Wechselwirkungen im zweifaktoriellen vollständigen Zufallsplan nicht möglich ist. Die optimistischen Erwartungen von Still & White (1981, S. 248), die auch für diesen Versuchsplan einen Permutationstest der alignierten Beobachtungen empfehlen und die Probleme von Wilk & Kempthorne (1955, S. 1153 f) für das finite Randomisierungsmodell unter wenig restriktiven Annahmen einen Wechselwirkungstest zu finden, werden deshalb im Licht des Ansatzes von Pyhel zu diskutieren sein.

Ein Hauptthema dieser Arbeit ist die Verallgemeinerung der Invarianz-Permutationstests auf multivariate Beobachtungen (Willmes, 1981a). Die Notwendigkeit von multivariaten Verfahren für die Psychologie wird in gängigen Lehrbüchern und Sammelreferaten zur multivariaten Analyse betont (vgl. Bock, 1975, Kap. 1.3; Morrison, 1976; Bray &

Maxwell, 1982; Moosbrugger & Steyer, 1983). Von allen Autoren wird angeführt, daß die simultane Berücksichtigung aller beobachteten Variablen einer Folge von univariaten Analysen überlegen ist, da sie mögliche Korrelationen der abhängigen Variablen berücksichtigt. Andererseits sind die stets notwendigen Voraussetzungen für die parametrischen Testverfahren wie mehrdimensionale Normalverteilung und Identität von Varianz-Kovarianz Matrizen unter verschiedenen experimentellen Bedingungen sehr restriktiv und gerade in den Verhaltenswissenschaften oft schwer zu rechtfertigen.

Über die Entwicklung der multivariaten Permutationstests hinausgehend wird auch zu untersuchen sein, welchen Stellenwert sie wie auch die univariaten Permutationstests in einer deduktivistischen Theorie des (psychologischen) Experimentes (Bredenkamp, 1980; Hager & Westermann, 1983a, b, c) haben können. Bredenkamp legt dar, daß - im Unterschied zu R.A. Fisher - es in einer deduktivistischen Theorie auch möglich sein muß, eine Entscheidung über die Gültigkeit der Nullhypothese zu treffen. Falls nämlich die wissenschaftliche (psychologische) Hypothese eine statistische (oft einseitige) Alternative impliziert, wäre diese ansonsten vor Falsifizierungen sicher. Es ist streng genommen mit einem Permutationstest nicht möglich, vor der Durchführung eines Experimentes den Mindeststichprobenumfang festzulegen, mit dem für eine gewünschte Teststärke ein bestimmter Mindesteffekt der experimentellen Bedingungen aufgedeckt werden soll. Im Gegensatz zum Niveau des Randomisierungstests hängt die Teststärke von den sog. "unit values" (Einheitenwerten; vgl. Kempthorne, 1952, 1955; Lehmann, 1975, S. 57) ab. Irgendeine Annahme über deren

Verteilung bzw. eine Annahme uber die Verteilung der Fehlervariablen im Modell von Pyhel ist also auch bei einem (bedingten) Permutationstest notwendig.

Eine andere Konsequenz der Forderung nach einer moglichst strengen Prüfung wissenschaftlicher Hypothesen mit statistischen Tests ist, daß nicht nur multiple Vergleiche oder Tests für einfache Haupteffekte im Anschluß an eine verworfene globale Nullhypothese verwendet werden, um Unterschiede zwischen verschiedenen Bedingungen aufzudecken, bzw. um eine Wechselwirkung zwischen unabhangigen Variablen näher zu bestimmen. Es ist also zu untersuchen, inwieweit die von Bredenkamp (1980, 1982) und Hager & Westermann (1983a, b, c) sowie Westermann & Hager (1983 a, b) vorgeschlagenen oder empfohlenen einseitigen Tests, Tests für monotone Trends und Tests zur Identifizierung einer ordinalen oder disordinalen Wechselwirkung auch mit Permutationstestverfahren ausgeführt werden können.

Obwohl, wie in den folgenden Abschnitten noch zu sehen sein wird, das Konstruktionsprinzip für Permutationstests sehr allgemein ist, wird sich die Arbeit auf das Testen von linearen Hypothesen in "verteilungsfreien" linearen Modellen mit uni- oder multivariaten abhängigen Variablen beschränken. Insbesondere wird immer die Beantwortung der Frage im Vordergrund stehen, für welche Tests in ANOVA - bzw. MANOVA-Modellen analoge Permutationstests konstruiert werden können. Lineare Hypothesen in linearen Modellen sind in der experimentellen (psychologischen) Forschung wohl die am häufigsten getesteten Hypothesen. Dagegen werden beim Einsatz von Rangtests, die in ihrer finiten Version oft

ebenfalls Permutationstests sind (vgl. Puri & Sen, 1971), häufig eigentlich größere Klassen von Alternativen wie z.B. sog. Tendenzhypothesen zugelassen (vgl. Brunner & Neumann, 1984, 1986; Erdfelder & Bredenkamp, 1984). Weiterhin ist in mehrfaktoriellen Plänen keine direkte Entsprechung zwischen den Haupt-und Wechselwirkungseffekten für die ANOVA und den formalen Analoga für Ränge mehr gegeben (siehe z.B. Lemmer, 1980; Hilgers, 1981; Erdfelder & Bredenkamp, 1984 und die problematischen Vorschläge von Gammel & Moosbrugger, 1982, oder Kubinger, 1986).

Verschiedene andere Fragestellungen, für die ebenfalls Permutationstests vorgeschlagen wurden, seien nur kurz aufgezählt:

Permutationstests für die Kovarianzanalyse werden von Cox (1956), Robinson (1973b), Peritz (1982), Roux (1982) und Bellhouse (1986) untersucht. Auch zur schrittweisen Variablenselektion in einer multiplen Regressionsanalyse wurde ein Permutationstest vorgeschlagen (Forsythe, Engelman, Jennrich & May, 1973). Weitere Permutationstests für regressionsanalytische Probleme stammen von Brown & Maritz (1982).

Der Vergleich von Verlaufskurven unter verschiedenen experimentellen Bedingungen wird von Zerbe & Walker (1977), Zerbe (1979) und Foutz, Jensen & Anderson (1985) behandelt. Die Teststatistiken basieren auf Summen quadrierter Abstande zwischen (mittleren) Verlaufskurven. Mit dem Vergleich von Überlebenszeiten unter verschiedenen Bedingungen beschäftigen sich Arbeiten von Cox & Kempthorne (1963) sowie Forsythe & Frey (1970). Marcuson & Nordbrook (1981) geben einen multivariaten Test für die Analyse von

Überlebenszeiten an. Von Peto (1973) stammt ein Programm für das univariate Problem.

Für kontrollierte klinische Studien mit sequentiell anfallenden Beobachtungseinheiten konnte eine zufällige Zuteilung jedes neuen Patienten zu einer von k Bedingungen (Therapien) mit der Wahrscheinlichkeit 1/k insbesondere bei kleinen bis mittleren Patientenanzahlen zu stark unterschiedlichen Stichprobenumfängen je Bedingung führen. Deshalb wurden verschiedene sog. restringierte Randomisierungsverfahren in Abhängigkeit von den bereits erfolgten Zuweisungen vorgeschlagen (Smythe & Wei, 1983; Wei, Smythe & Smith, 1986) oder diskutiert (Cox, 1982).

Auch die Vorschläge zur Analyse von Einzelfallstudien mit Randomisierungstests, in denen verschiedene Interventionen verglichen werden, sei nur erwähnt, obwohl verschiedene in Abschnitt 3.1.5 beschriebene Teststatistiken auch für diesen Fall (N = 1) verwendet werden können (Edgington, 1975 b, 1980, Kap. 10; Levin, Marascuilo & Hubert, 1978; Holtz & Eberle, 1978; Möbus & Nagl, 1983; Dahme, 1984).

Nicht unerwähnt bleiben soll, daß Fisher (1935, Kap. 2) selbst als erstes Beispiel für die Anwendung eines Randomisierungstests ein fiktives Einzelfallexperiment ("The Lady tasting tea experiment"; Gridgeman, 1959; Box, 1978; Foutz, 1984) betrachtet hat.

Schließlich gibt es beginnend mit Pitman (1937b) eine größere Anzahl von Arbeiten, die sich mit Permutationstests zur Analyse der Unkorreliertheit von zwei Merkmalen beschäftigen (Daniels, 1944; Edgington, 1980). Während Pitman wie auch van Elteren & Verbeek (1979) nur den Produktmoment-Korrelationskoeffizienten betrachten, läßt

sich bei Betrachtung von sog. verallgemeinerten (partiellen) Assoziationsmaßen eine Vielzahl von Fragestellungen in einheitlicher Weise behandeln wie der Vergleich von (Ähnlichkeits-)Matrizen (Hubert & Schultz, 1968; Hubert, 1985), der Vergleich von soziometrischen Wahlen (Hubert & Baker, 1978), die Analyse von Beurteilerübereinstimmungen (Hubert & Golledge, 1983). Auch die Analyse von raum-zeitlicher Häufungen eines Merkmals (z.B. von infektiosen Krankheiten) läßt sich in ähnlicher Weise analysieren (Mantel, 1967; Mantel & Valand, 1970; Klauber, 1971, 1975; Tsutakawa & Yang, 1974; Besag & Diggle, 1977).

Abschließend soll eine Übersicht über den Aufbau der Arbeit gegeben werden:

In Abschnitt 2 werden ohne notwendigen Bezug auf ein lineares Modell das Konstruktionsprinzip von Permutations- und Randomisierungstests und deren wichtigste testtheoretische Eigenschaften erlautert sowie einige Grundbegriffe eingeführt.

Die Frage der Existenz von Invarianzpermutationstests für lineare Hypothesen im "verteilungsfreien" linearen Modell bildet den Schwerpunkt von Abschnitt 3.

Als verteilungsfrei werden solche linearen Modelle charakterisiert, in denen die Fehlerkomponenten nicht unabhangig und identisch normalverteilt sind, sondern in denen die gemeinsame Verteilung der Fehler lediglich gewissen Vertauschbarkeits- d.h. Symmetrieeigenschaften genugen muß. Diese vorausgesetzten Symmetrieeigenschaften sind oft durch die im Experiment vorgenommene zufallige Zuteilung der Beobachtungseinheiten zu den experimentellen Bedingungen motiviert und umfassen ganze Klassen von

Verteilungen.

Der von Pyhel (1978a) für Permutationstests entwickelte Ansatz, durch das jeweilige Testproblem invariant lassende Transformationen des Parameterraumes Tests für einen "Teil der Parameter" im linearen Modell unbeeinflußt von den übrigen "störenden" Parametern zu ermoglichen, wird ausfuhrlich erläutert. Von entscheidender Bedeutung für die Existenz eines Permutationstests ist, welche Vertauschbarkeitseigenschaften die Fehlerverteilung nach der Ausführung der Invarianz des Testproblems sichernden Transformationen noch besitzt, um einen nichttrivialen Permutationstest konstruieren zu können. Für die meisten in der experimentellen psychologischen Forschung üblichen ein- bis dreifaktoriellen Versuchspläne wird die Existenz von Invarianzpermutationstests für Haupt- und Wechselwirkungsparameter untersucht und - wenn möglich - ein solcher Test angegeben. Insbesondere wird die von Pyhel (1978a, b) nur kurz erwähnte Nichtexistenz eines Wechselwirkungstests im zweifaktoriellen vollstandigen Zufallsplan ausführlich dargestellt (vgl. auch Erdfelder & Bredenkamp, 1984, S. 276). Neu sind die Ergebnisse für Tests bezuglich der dreifachen Wechselwirkung in dreifaktoriellen Plänen. Es wird sich zeigen, daß nur im dreifaktoriellen Blockplan und im Split-Split-Plot Plan ein Invarianzpermutationstest angegeben werden kann. Je nach Vertauschbarkeitsforderungen an die gemeinsame Fehlerverteilung werden auch verschiedene Tests für (verallgemeinerte) Profilparallelität vorgestellt. Die von Pyhel (1980a) für den Split-Plot Plan entwickelten Tests auf Profilparallelität für kleine und große Stichprobenumfänge

werden damit verallgemeinert. Die Teststatistiken haben wiederum asymptotisch (unbedingte) Chi-Quadrat Verteilungen. Die vorgeschlagenen Tests geben eine für bestimmte Versuchspläne positive Antwort auf die von Erdfelder & Bredenkamp (1984, S. 276) geäußerte negative Erwartung bezüglich der Existenz von Wechselwirkungstests für faktorielle Blockpläne.

Das Prinzip der invarianten Reduktion eines Testproblems kann auch für die Analyse der Existenz von Randomisierungstests in linearen Modellen genutzt werden. In Abschnitt 3.3 wird gezeigt, daß auch die diskrete gemeinsame Verteilung der sog. Design-Zufallsvariablen, die die zufällige Zuteilung der Beobachtungseinheiten zu den experimentellen Bedingungen repräsentieren, unter Invarianz sichernden Transformationen unter Umständen so wenig an Symmetrieeigenschaften übrig behält, daß ein exakter Invarianzpermutationstest für dieselben linearen Hypothesen wie in Abschnitt 3.1 im Randomisierungsmodell ebenfalls nicht existiert, auch wenn approximativ F-verteilte Teststatistiken mit korrigierten Freiheitsgraden von Kempthorne und Mitarbeitern sowie anderen Autoren vorgeschlagen worden sind.

Weitgehend neu ist die Verallgemeinerung der Invarianzpermutationstests auf multivariate Probleme in Abschnitt 4. Es wird gezeigt, daß eine invariante Reduktion des Testproblems wiederum möglich ist. Auch die Permutationsinvarianzstrukturen der transformierten Fehlervektoren sind zu denen im univariaten Fall isomorph. Es sind nun lediglich (d-dimensionale) Beobachtungsvektoren als ganze zu permutieren. Die Struktur der unter der

jeweiligen multivariaten Nullhypothese zulassigen (Vektor-) Permutationen ist zu allen in Abschnitt 3 betrachteten Testproblemen analog. Zu diskutieren ist jedoch die Wahl einer geeigneten Teststatistik, die sich wie im MANOVA-Fall als Eigenwerte gewisser Determinantengleichungen darstellen lassen. Das schon von Mardia (1971) verwendete Spurkriterium von Pillai stellt sich wiederum als auch vom Rechenaufwand besonders vertretbar heraus. Verteilungsfreie sog. Profilanalysen (Timm, 1980) sind mit dem Permutationstestansatz ebenfalls möglich. Diese werden in Abschnitt 4.6 behandelt. Abschließend wird demonstriert, daß für mehrfaktorielle Pläne Tests von Wechselwirkungsparametern mit den sog. MRPP-Statistiken (Mielke, Berry & Johnson, 1976) konstruierbar sind. Diese Statistiken sind aus Summen normierter Distanzen zwischen (invariant transformierten) Beobachtungsvektoren gebildet.

Permutationstests für einige multiple Testprobleme werden in Abschnitt 5 vorgestellt. Es wird vornehmlich nur der Fall des Vergleichs aller Paare von Faktorstufen betrachtet. Es wird sich zeigen, daß die Arbeiten von Sonnemann (1981a, b, 1982) zu Tests zum multiplen Niveau alpha einen geeigneten konzeptuellen Rahmen auch für Permutationstests abgeben (Abschnitt 5.2). Insbesondere sequentiell verwerfende Tests nach Holm (1979) und Marcus, Peritz & Gabriel (1976) sind auf die Struktur von multiplen Permutationsinvarianz-hypothesen bezüglich des Vergleichs von Faktorstufen zugeschnitten, wie bereits Petrondas & Gabriel (1983) verdeutlicht haben (Abschnitt 5.4).

Auch für die Komponenten multivariater Beobachtungen lassen sich exakte sequentielle multiple Permutationstests angeben. Ein Vorschlag von Maurer (1981) wird aufgegriffen und seine generelle Anwendbarkeit auch für generellere Hypothesenfamilien demonstriert (Abschnitt 5.4 und 5.6). Multiple Permutationstests, die lediglich das globale Niveau alpha kontrollieren, sind von Foutz (1981) zuerst vorgeschlagen worden. Sie werden in Abschnitt 5.5 behandelt.

Die beiden folgenden Abschnitte sind besonders für die Testplanung und eine deduktivistische Theorie des Experimentierens (Bredenkamp, 1980), sofern sie sich auf die Verwendung von Signifikanztests bezieht, von Bedeutung.

Abschnitt 6 beginnt mit prinzipiellen Problemen bei der Bestimmung der Teststärke von bedingten Permutationstests. Nur asymptotisch läßt sich eine unbedingte Teststärke für Permutationstests angeben, die identisch mit dem ANOVA-Fall ist, sofern Normalverteilung vorliegt. Im finiten Fall läßt sich eine exakte konditionale Teststärke zwar definieren (Gabriel & Hall, 1983), aber wegen des immensen Rechenaufwandes praktisch nicht bestimmen. Deshalb werden verschiedene Vorschläge zur approximativen Bestimmung der Teststärke eines Permutationstests nach Ausführung eines Experimentes gemacht (sog. aposteriori Teststärkeschätzung). Für die Testplanung bleibt neben der in Abschnitt 7 betrachteten Näherungslösung mit Hilfe des Normalverteilungsfalles nur die apriori Schätzung der sog. naiven empirischen Teststärke in Simulationsexperimenten mit Hilfe von Stichproben aus empirischen Residuenverteilungen gewonnen in vorhergehenden Experimenten oder die Generierung von Zufallsstichproben aus bestimmten Residuenverteilungen. In diesem Zusammenhang ist auch die Robustheit von

Permutationstests (Abschnitt 6.4) von Interesse. In dieser Arbeit werden jedoch keine Ergebnisse eigener Robustheitsstudien vorgelegt, sondern nur eine Literaturübersicht gegeben.

In Abschnitt 7 wird ausführlich untersucht, in welchem Umfang die bisher vorgestellten und entwickelten Permutationstestverfahren von Bedeutung für eine deduktivistische Theorie des Experimentes sind. Da in der experimentellen psychologischen Forschung in aller Regel keine Zufallsstichproben aus sinnvoll definierbaren Populationen vorliegen, sind Testverfahren, die sich konzeptuell an der experimentellen Randomisierung "orientieren", wünschenswert. Wenn man bei der Planung des in allen Experimenten benötigten Stichprobenumfanges mit approximativen Lösungen zufrieden ist, für die man gute Gründe anführen kann, (Abschnitt 7.3), sind Permutationstests den üblichen (M)ANOVA-Verfahren überlegen. Zumal in Abschnitt 7.4 gezeigt wird, daß auch die möglichst strenge Prüfung wissenschaftlicher Hypothesen mit Hilfe von Permutationstests in vielen Fällen möglich ist. Einseitige uni-und multivariate Permutationstests sind oft möglich. Über die bei Edgington (1980) zu findenden Vorschläge für das Testen von (monotonen) Trendhypothesen in Permutationstests hinaus wird die Übertragbarkeit vieler parametrischer Verfahren auch für mehrfaktorielle Versuchspläne aufgezeigt. Den Abschluß der Arbeit bilden Vorschlage, mit Permutationstests den Typ (ordinal bzw. disordinal) einer Wechselwirkung im Sinne von Bredenkamp (1982) zu bestimmen.

Alle soweit entwickelten Permutationstests sind nur dann praktisch nutzbar, wenn man handhabbare Computerprogramme zu ihrer Ausführung zur Verfügung hat. Denn selbst für kleine Stichprobenumfänge ist die Bestimmung der Überschreitungswahrscheinlichkeit (p-Wert) ohne Computer kaum möglich. Es muß ja für jede unter der jeweiligen Nullhypothese zulässige Permutation der Beobachtungen ein Wert der jeweiligen Teststatistik berechnet werden. Im Abschnitt 8 wird nach einem Überblick über bereits publizierte Verfahren die sog. Monte-Carlo Version des Permutationstests diskutiert. Gestützt auf die theoretischen Arbeiten von Dwass (1957) und besonders Jöckel (1982, 1986a, b) bietet sie die Möglichkeit, auch für größere Stichprobenumfänge Tests mit einer nur geringen Einbuße an Teststärke verglichen mit der exakten Version des Permutationstests ausführen zu können. Nach der Illustration des allgemeinen Aufbaus von exakten und Monte-Carlo-Permutationstests anhand von Flußdiagrammen, werden spezielle FORTRAN -Algorithmen für die Ausfuhrung uni-und multivariater Permutationstests in Weiterentwicklung der Programme von Pyhel (ohne Jahr) beschrieben. Ein Programm von Petrondas & Gabriel (1981) wird in allgemeinerer Form für sequentielle multiple uni- und multivariate Testprobleme dargestellt. Neu ist ein Algorithmus für die sequentielle multiple Testprozedur bezüglich Komponenten multivariater Beobachtungen (Abschnitt 8.4.4), die vielseitig einsetzbar ist. Die Computerprogramme werden beispielhaft bei der Reanalyse von publizierten Datensätzen oder bei der Analyse von Daten aus eigenen Untersuchungen eingesetzt. Im Anhang sind beispielhaft wesentliche Bestandteile der eigenen Programme aufgelistet.

2. DAS PRINZIP DES PERMUTATIONSTESTS UND DES RANDOMISIERUNGSTESTS

Die konzeptuellen Unterschiede zwischen einem Permutations- und einem Randomisierungstest werden besonders von Kempthorne (1975, 1977) und Kempthorne & Doerfler (1969) herausgearbeitet.

Mit einem Permutationstest werden Zufallsstichproben aus endlichen oder unendlich großen Grundgesamtheiten analysiert oder verglichen. Die Verteilungsfunktionen der Zufallsvariablen, die diese Zufallsstichproben beschreiben, sind nicht näher spezifiziert und müssen lediglich aus einer großen Klasse von Verteilungsfunktionen stammen, z.B. aus der Klasse aller nichtdegenerierten oder aller stetigen Verteilungsfunktionen. Die Validität eines Permutationstests bezüglich Aussagen über die interessierenden Grundgesamtheiten hängt (allein) davon ab, daß die analysierten Stichproben Zufallsstichproben sind.

Ein Zufallsexperiment mit n Beobachtungen kann man durch eine n-dimensionale Zufallsvariable $X=(X_1,\ldots,X_n)'$ beschreiben, die nach der n-dimensionaler Verteilungsfunktion F(x) verteilt sei. Das Ergebnis eines Experimentes ist dann eine Realisation $x=(x_1,\ldots,x_n)'\in\mathbb{R}^n$. Mit Permutationstests können nun solche Hypothesen getestet werden, die annehmen, daß die gemeinsame Verteilung F(x) der Beobachtungen bei Gültigkeit der betreffenden Hypothese invariant (unverändert) bleibt unter bestimmten Arten von Vertauschungen (Gruppen von Permutationen). F(x) muß also bestimmte Symmetrieeigenschaften besitzen. Sei S_n die Menge (genauer Gruppe) aller n! möglichen Permutationen von n Objekten. Alle üblichen Hypothesen in Permutationstests mit Ausnahme

der Vorzeicheninvarianzhypothese lassen sich nun als Invarianzen von F(x) gegenüber bestimmten Teilmengen $\mathcal{Q}_n$ von S_n (genauer: Untergruppen von S_n) darstellen (vgl. Scheffe, 1959. S. 317). Die Anzahl der Permutationen in $\mathcal{Q}_n$ sei $M \leq n!$. Die Nullhypothese lautet dann:

$$(2.1) \qquad H^P: \quad F(x) = F(\pi x) \quad \text{für alle } \pi \in \mathcal{Q}_n \text{ und alle } x \in R^n .$$

Die Vorzeicheninvarianzhypothese beinhaltet dagegen, daß die Verteilung von X invariant bleibt unter allen $M=2^n$ möglichen Vorzeichenvertauschungen der n Komponenten von X.
Nun betrachtet man eine Funktion der n Beobachtungen S(X), die Teststatistik, die jede Realisation der Zufallsvariablen X in eine reelle Zahl abbildet. Sei die Teststatistik so gewahlt, daß sie große Werte annimmt, wenn die Nullhypothese H^P nicht zutrifft. Unter der Nullhypothese sind dann alle Werte $s_1,\ldots,s_n$, die S(X) für die Permutationen aus $\mathcal{Q}_n$ annimmt, gleichwahrscheinlich mit der Wahrscheinlichkeit 1/M. Man kann nun die Werte der Teststatistik der Große nach ordnen zu $s_{(1)} \leq \ldots \leq s_{(M)}$. Die $s_{(i)}$ brauchen nicht notwendig verschieden zu sein. Seien $K \leq M$ der Teststatistikwerte verschieden und trete jeder von ihnen k_l-mal auf. Sei $k_l \geq 1$ und $\sum_{l=1}^{K} k_l = M$. Dann kann man die (empirische) Permutationsverteilungsfunktion der Teststatistik, eine 'Treppenfunktion' folgendermaßen darstellen, wobei noch angenommen sei, daß $s_{(1)} > 0$ gelte:

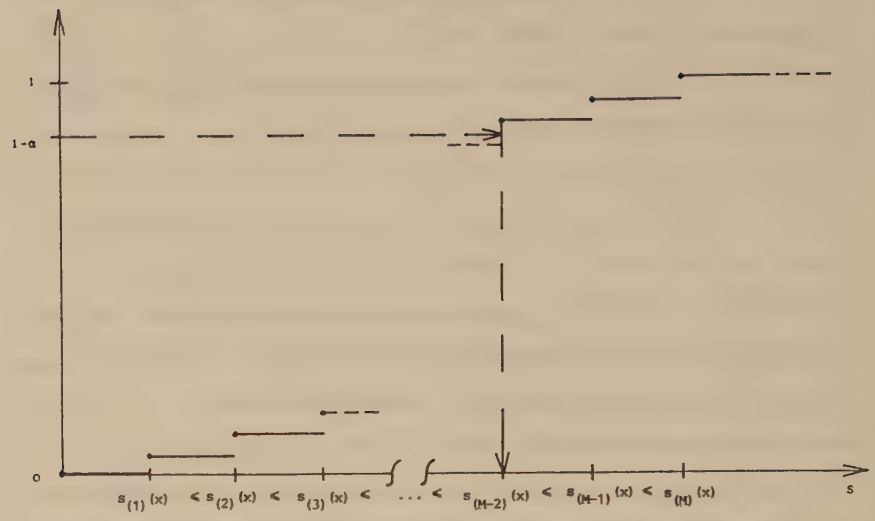


Abb. 2.1.: Permutationsverteilungsfunktion der Teststatistik S(X) und Bestimmung des Schwellenwertes

Cox & Hinckley (1974) wie auch Kempthorne (1976) definieren nun einen Signifikanztest als die Ermittelung des (empirischen) Signifikanzniveaus p - oft auch als p-Wert bezeichnet - bei beobachtetem Wert s(x) der Teststatistik S(X) als:

$$(2.2) \qquad p := p(x) = P(S(X) \geq s(x) \mid H^P) \quad , 0 \leq p \leq 1 \ ,$$

d.h. als Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese Werte der Teststatistik zu beobachten, die größer oder gleich dem beobachteten Wert sind. Damit ist p(X) selbst eine Teststatistik. Der Statistikwert p wird als Maß für die Konsistenz der beobachteten Daten mit der Nullhypothese interpretiert. Je kleiner p ist, umso größer ist die

'Evidenz gegen H^p' (vgl. Gibbons & Pratt, 1975).

Im Sinne der Testtheorie von Neyman & Pearson (1933a, b; vgl. z.B. Lehmann, 1959) werden Tests durch Entscheidungsfunktionen $\phi(x), x\in R^n$ mit $0\leq\phi(x)\leq 1$ dargestellt. ϕ nimmt bei Verwerfen der Nullhypothese den Wert 1 und bei Nichtverwerfen den Wert 0 an. Falls $\phi(x)=\gamma$, $0<\gamma<1$, ist, trifft man mit Hilfe eines zusatzlichen Zufallsexperimentes eine 'randomisierte' Entscheidung, d.h. die Nullhypothese wird mit Wahrscheinlichkeit γ verworfen. Auf eine mathematisch exakte Darstellung der Permutationstests, die die Theorie der bedingten Tests benotigt,sowie eine Klärung der Meßbarkeitseigenschaften wird nicht eingegangen (vgl. Witting & Nölle, 1970; Pyhel, 1978a).

Die Entscheidungsregel für einen (einseitigen) Permutationstest zum Niveau α (für den Fehler 1. Art), $0<\alpha<1$, läßt sich nun folgendermaßen formulieren:

Sei $[M\alpha]$ die größte ganze Zahl, die kleiner oder gleich $M\alpha$ ist und sei $m:M-[M\alpha]$. Dann bildet man:

$$(2.3)\qquad \begin{aligned} p^{>} &:= p^{>}(x) := |\{s_{(j)}(x) \mid s_{(j)}(x) > s(x)\}|/M \\ p^{=} &:= p^{=}(x) := |\{s_{(j)}(x) \mid s_{(j)}(x) = s(x)\}|/M \quad . \end{aligned}$$

Der Wert $s_{(m)}(x)$ heißt auch m-Schwellenwert der Permutationsverteilungsfunktion von S. Mit der Wahl $\gamma=(\alpha-p^{>}(x))/p^{=}(x)$ lautet der Test (die Entscheidungsregel) zum Niveau α:

$$\phi(x) = \begin{cases} 1 & S(x) > s_{(m)}(x) \\ \gamma & S(x) = s_{(m)}(x) \\ 0 & S(x) < s_{(m)}(x) \end{cases}$$

Man kann sich leicht uberlegen (vgl. Pyhel, 1980a), daß man zur Durchführung des Tests die möglichen Werte der Teststatistik nicht zu ordnen braucht, was sehr aufwendig wäre. Vielmehr genügt es, als äquivalente Entscheidungsregel

$$(2.4) \qquad \phi(x) = \begin{cases} 1 & p^{>} + p^{=} \leq \alpha \\ \gamma & \alpha - p^{=} < p^{>} \leq \alpha \\ 0 & p^{>} > \alpha \end{cases}$$

zu betrachten. Es müssen also nur die relativen Anteile $p^{>}$ und $p^{=}$ ermittelt werden.

Obiger Test ist für eine einseitige Entscheidung geeignet, denn nur 'große' Werte der Teststatistik deuten auf eine Verletzung der Nullhypothese hin. Für ein zweiseitiges Testproblem findet man analog:

$$(2.5) \qquad \phi(x) = \begin{cases} 1 & p^{>} + p^{=} \leq \alpha/2 \\ \gamma_1 & \alpha/2 - p^{=} < p^{>} \leq \alpha/2 \\ 0 & \alpha/2 < p^{>} < 1 - \alpha/2 - p^{=} \\ \gamma_2 & 1 - \alpha/2 - p^{=} \leq p^{>} < 1 - \alpha/2 \\ 1 & 1 - \alpha/2 \leq p^{>} \end{cases}$$

Alle nach diesem Prinzip erstellten Permutationstests zum Testen einer Invarianzhypothese haben per Konstruktion die Eigenschaft, exakte (ähnliche) Tests zum Niveau α zu sein (vgl. Scheffe, 1943b);Hoeffding, 1952; Lehmann, 1959).

Wenn man auf die randomisierte Entscheidung in (2.4) oder (2.5) verzichtet, wofür es vernünftige Gründe gibt (vgl. Scheffe, 1959, S. 317; Kempthorne & Doerfler, 1969, S. 234), hat man als Entscheidungsfunktion:

$$\phi'(x) = \begin{cases} 1 & p^{>} + p^{=} \leq \alpha \\ 0 & \text{sonst} \end{cases}$$

Dieser Test hat ein Niveau kleiner oder gleich α . Der nichtrandomisierte Permutationstest kann somit nicht jedes beliebige Niveau annehmen, sondern nur ganzzahlige Vielfache von $1/M$.

Ohne Ruckgriff auf Entscheidungsfunktionen lassen sich die obigen Entscheidungsregeln auch anschaulicher folgendermaßen formulieren (vgl. König, Giani & Haux, 1986). Man betrachtet

den nichtrandomisierten p-Wert

$$p = p^{>} + p^{=}$$

(2.7) oder den randomisierten p-Wert

$$p = p^{>} + u \cdot p^{=} \quad .$$

Dabei ist u eine Realisation einer auf dem Intervall [0,1] gleichverteilten Zufallsvariablen. Es wird also stets nur ein bestimmter Anteil der mit dem beobachteten Wert der Teststatistik identischen Teststatistikwerte bei permutierten Beobachtungen zur Bildung des p-Wertes herangezogen.
Die Nullhypothese H^P der Permutationsinvarianz wird verworfen, falls p nicht größer als das festgesetzte Niveau α ausfällt. Gelegentlich ist auch empfohlen worden (Franck, 1986), den konstanten Wert u=0.5 zu wählen (sog. 'mid-p value'). Der randomisierte p-Wert wird lediglich benutzt, um eine möglichst große Teststärke des Permutationstests zu erzielen. Der nichtrandomisierte p-Wert ist deshalb 'konservativ'.
Für zweiseitig formulierte Hypothesen, bei denen 'sehr große' und 'sehr kleine' Werte der Teststatistik gegen die Nullhypothese sprechen, sind mehrere Vorschlage zur Definition eines zweiseitigen p-Wertes gemacht worden. Üblich ist die folgende Version:

(2.8) $$p = 2 \cdot \min\{p^{>}+p^{=},\ p^{<}+p^{=},\ 0.5\} \quad .$$

Der Wert 0.5 wird mit angeführt, um zu verhindern, daß p großer als 1 werden kann. Gibbons & Pratt (1975) schlagen für nichtrandomisierte Entscheidungen bei zweiseitigen Testproblemen vor, die Richtung der Abweichung von der Null-

hypothese zusammen mit dem zugehörigen einseitigen p-Wert anzugeben.

Permutationstests sind bedingte Tests, d.h. die Testfunktion ist abhängig von der jeweiligen Realisation der Zufallsvariablen X. Damit sind keine Vertafelungen von Schwellenwerten möglich. Für jedes Experiment müssen erneut $p^{>}$ und $p^{=}$ bestimmt werden. Die Anzahl der Teststatistikwerte steigt mit wachsendem Stichprobenumfang n sehr schnell an (vgl. Streitberg, 1986).

Abhilfe schafft die 'Monte Carlo' Version des Permutationstests oder Approximationen der Permutationsverteilung der Teststatistik (vgl. Abschnitt 8).

Das Prinzip des Randomisierungstests läßt sich folgendermaßen verdeutlichen. Man betrachtet n Beobachtungseinheiten und die Menge aller möglichen Versuchsschemata, mit denen die n Einheiten einer von $k \geq 2$ experimentellen Bedingungen zugewiesen werden können. Alle diese M Versuchsschemata seien so beschaffen, daß stets dieselbe Anzahl von Einheiten je einer der Bedingungen zugewiesen wird. Dann wird zufällig ein Versuchsschema ausgewählt und ein Beobachtungswert x_i, $i=1,..,n$, für jede der Beobachtungseinheiten registriert. Die k experimentellen Bedingungen sollen nun verglichen werden. Für diesen Vergleich ist es unerheblich, ob die Beobachtungseinheiten zufällig oder systematisch aus einer endlichen oder unendlichen Grundgesamtheit ausgewählt worden sind oder ob gar kein Bezug auf eine Grundgesamtheit genommen wird. Für den Vergleich ist es nur wichtig, daß der 'physikalische Akt' der zufalligen Auswahl eines Planes bzw. der zufälligen Zuteilung der Einheiten zu den Bedingungen

erfolgreich ausgeführt wurde (Lehmann, 1959, S. 282; Bredenkamp, 1980). Die in diesem Fall zu betrachtende Grundgesamtheit ist die Grundgesamtheit aller möglichen oder zulässigen vom Experimentator ausgewählten Versuchspläne. Es ist nicht notwendig, daß alle Pläne mit der gleichen Wahrscheinlichkeit 1/M ausgewählt werden (können) (vgl. Cox, 1956; Rosenbaum, 1984). Diese Bedingung wird jedoch nachfolgend als erfüllt angesehen. Die Grundgesamtheit von möglichen Wiederholungen des Experimentes ist erst nach der zufälligen Auswahl einer bestimmten Anordnung der k Versuchsbedingungen eine lediglich konzeptuelle Grundgesamtheit.

Die übliche (globale) Nullhypothese in einem Randomisierungstest lautet:

(2.9) H^R: Alle k Versuchsbedingungen haben für alle betrachteten n Einheiten denselben Effekt.

Für jeden aus den M Plänen zufällig ausgewählten Versuchsplan kann sich eine Variation in den Komponenten des Beobachtungs- oder Meßwertvektors $x= (x_1,..,x_n)'$ nur durch Variabilität in den Beobachtungseinheiten selbst ergeben, wenn man Meßfehler vorerst unberücksichtigt läßt. Die Einflüsse der Einheiten werden in den sog. 'unit values' (Einheitenwerten) oder 'basic yields' $u_i, i=1,..,n$, erfaßt. In der Agrarforschung, für die diese Modelle zuerst entwickelt wurden, sind das beispielsweise die Erträge eines Stückes Land ohne die Anwendung eines von k zu vergleichenden Düngemitteln. In einem psychophysiologischen Experiment konnen es die unter einer Ruhebedingung gewonnenen Hautwiderstands- oder Blutdruckwerte sein.

Die Zufallszuteilung zu den experimentellen Bedingungen läßt sich durch (diskrete) Zufallsvariablen, die sog. 'Designvariablen' darstellen:

$$\delta_i^j = \begin{cases} 1 & \text{i-te Einheit wird j-ter Bedingung} \\ & \text{zugewiesen, } j=1,..,k \\ 0 & \text{sonst} \end{cases}$$

Wie zuvor wählt man für eine Testentscheidung eine 'geeignete' Teststatistik S, die 'große' Werte annehmen soll, wenn die verschiedenen experimentellen Bedingungen tatsächlich unterschiedliche Effekte auf die Beobachtungseinheiten ausüben. Die Teststatistik $S(x_1,..,x_n)$ ist eine Funktion der Designvariablen und der Einheitenwerte $u=(u_1,...,u_n)'$. Wie zuvor in (2.2) kann man wieder zur Testentscheidung den p-Wert benutzen:

$$\begin{aligned} p := p(x) &= P(S(x,\{\delta_i^j\}) \geq s \,|\, H^R \text{ und } u) \\ &= 1/M \cdot |\{\{\delta_i^j\}: S(x,\{\delta_i^j\}) \geq s\}| \quad ; \end{aligned} \tag{2.10}$$

s sei der Wert der Teststatistik für die tatsächlich beobachteten Werte.

Die nichtrandomisierte Testentscheidung lautet dann für eine einseitige Hypothese zu einer vorgegebenen Irrtumswahrscheinlichkeit $\alpha, 0<\alpha<1$, wieder:

(2.11) H^R wird verworfen, falls $p \leq \alpha$ ist.

Offensichtlich gilt dann (vgl. Kempthorne & Doerfler, 1969, S. 233):

(2.12) $P(p(x) \leq \alpha \,|\, H^R \text{ und } u) \leq \alpha$.

Gleichheit ist in (2.12) dann gegeben, wenn α ein 'erreichbares' Fehlerniveau ist.

In Analogie zum Permutationstest sind wiederum ein zweiseitiger Test und ein randomisierter p-Wert definierbar.

Hervorgehoben sei noch einmal, daß ohne jede Verteilungsannahme für die u_i die Beziehung (2.12) gültig ist. Wegen dieser Abhängigkeit von den u_i ist auch für Teststatistiken von Randomisierungstests keine Vertafelung von Schwellenwerten (kritischen Werten) moglich. Aus demselben Grund ist die Wahl einer optimalen (möglichst teststarken) Teststatistik nicht zu verwirklichen.

Wie zuvor angedeutet, ist es möglich, Meßfehler in das Modell mit einzubeziehen; darauf wird in Abschnitt 3.3 eingegangen. Jede der n Bedingungen kann auch selbst mehrdimensional sein. Die Menge der Versuchspläne, aus denen eine Versuchsanordnung zufällig ausgewählt wird, kann erweitert werden, wenn man die Bedingung aufgibt, daß in jedem der Pläne jede der Bedingungen mit der gleichen Haufigkeit auftritt (vgl. Hartigan, 1969; Gabriel & Hall, 1983).

Eine Diskussion der beiden Arten von Tests mit ihren Implikationen für eine deduktivistische Theorie des Experimentierens erfolgt in Abschnitt 7.

3. PERMUTATIONS- UND RANDOMISIERUNGSTESTS FÜR LINEARE HYPOTHESEN IN LINEAREN MODELLEN

Ein wichtiges Problem der Inferenzstatistik ist es, verteilungsfreie Tests für Hypothesen bezüglich mehrfaktorieller Versuchspläne zu konstruieren. Insbesondere sind Tests für lineare Hypothesen in mehrfaktoriellen linearen Modellen ohne die üblichen Normalverteilungsannahmen des klassischen linearen Modells (s. z.B. Kirk, 1968; Winer, 1971) von Interesse. Die Abschwächung der Verteilungsannahmen unter Beibehaltung der Struktur der Modellparameter führt aber schon zu Komplikationen. Die Hauptforderung an den Test eines Faktoreffekts (Haupt- oder Wechselwirkung), seine Entscheidung unabhängig von den anderen auftretenden Effekten zu treffen, ist nur unter bestimmten Bedingungen erfüllbar. Für mehrfaktorielle "Rangvarianzanalysen" sind seit der Arbeit von Lemmer & Stoker (1967) über den zweifaktoriellen Zufallsplan (CRF - pq in der Notation von Kirk, 1968) verschiedene Tests vorgeschlagen worden, die diese Forderung der Unabhängigkeit nicht beachten. Eine Übersicht und Kritik dieser Verfahren geben z.B. Erdfelder & Bredenkamp (1984). Es gibt jedoch auch einige Rangverfahren für spezielle (lineare) Hypothesen von Mehra & Sarangi (1967); Mehra & Sen (1969); Puri & Sen (1967); Sen (1968); Shiraishi (1984), die die obige Forderung erfüllen. In diesen 'ranking after alignment'-Verfahren werden die nicht zu testenden, 'störenden' Effekte durch spezielle Transformationen der Beobachtungen vor Vergabe der Ränge 'entfernt'.

Diese Tests sind, wenn man die abschließende Rangtransformation nicht berücksichtigt, Spezialfälle des allgemeine-

ren Ansatzes von Pyhel (1978, 1980a, b). Da sein Vorgehen für die Konstruktion von (Invarianz-) Permutationstests in der uni-und multivariaten Datenanalyse grundlegend ist, wird er im folgenden Abschnitt 3.1 ausführlich dargestellt. In Abschnitt 3.2 wird das Prinzip des Invarianzpermutationstests diskutiert. Abschließend kann in Abschnitt 3.3 gezeigt werden, daß unter Invarianzforderungen auch im Randomisierungsmodell für bestimmte Testprobleme wie im Permutationsmodell kein Permutationstest existiert.

3.1 Invarianz- Permutationstests

3.1.1 Reduktion eines Testproblems durch Invarianz

Ausgangspunkt ist das lineare Modell

(3.1) $$Y = Xb + \varepsilon \quad .$$

Die n-dimensionale Zufallsvariable Y beschreibt den Vektor der (univariaten) Beobachtungen. X ist die (nxk)-dimensionale Designmatrix vom (vollen) Rang k, wie sie üblicherweise zur Beschreibung eines linearen Modells verwendet wird (Bock, 1975; Moosbrugger & Steyer, 1983), in dem schon gewisse Reparametrisierungsbedingungen für den k-dimensionalen Parametervektor b mit $k<n$ erfüllt seien. Für die n-dimensionale Verteilungsfunktion von Y gelte, daß sie durch den n-dimensionalen Lageparameter $\theta=Xb$ charakterisiert wird. Man schreibt

(3.2) $$F_\theta(y) = F(y - \theta) \quad , \quad \theta \in U \subset R^n \quad .$$

θ sei Element eines linearen Unterraumes U des R^n mit der Dimension k. Für F sei weiterhin nur gefordert, daß gewisse

Symmetrieeigenschaften gelten:

(3.3) $F(x) = F(\pi x)$ für alle $\pi \in \mathfrak{Q}^*_n$ und $x \in R^n$.

$\mathfrak{Q}^*_n$ sei wie in (2.1) eine Permutationsgruppe.

Das lineare Modell (3.1) mit der 'schwachen' Forderung (3.3) an die Fehlerverteilung wird von Pyhel als 'verteilungsunabhängiges lineares Modell' bezeichnet. Diese Annahmen werden in Abschnitt 3.1.3 näher erläutert.

Lineare Hypothesen in dem linearen Modell lassen sich so formulieren:

(3.4) $H: \quad \theta \in U^H$,

d.h. $\theta \in R^n$ ist Element eines linearen Unterraumes U^H von U mit der Dimension h, $h \leq k$, gegen die Alternative

$$A: \quad \theta \in U \setminus U^H$$

($U \setminus U^H$ ist kein linearer Unterraum des R^n, da das Nullelement 0 nicht enthalten ist).

Das Testproblem (3.1) bzw. (3.4) bleibt invariant unter allen Transformationen $g: R^n \to R^n$ der Stichproben-Zufallsvariable Y von der Art:

$$g = id + \theta^* \,, \quad \theta^* \in U^H \quad .$$

Diese Transformationen bilden eine Gruppe $G=\{g \mid id+\theta, \theta \in U^H\}$ (vgl. Lehmann, 1959, Kap. 6.1)

Da U^H ein linearer Teilraum ist, ist mit

$$g(Y) = \theta + \varepsilon + \theta^* = \theta + \theta^* + \varepsilon$$

auch $\theta + \theta^* \in U^H$, falls $\theta, \theta^* \in U^H$.

Sei $f_H: R^n \to R^n$ die orthogonale Projektion des R^n auf den linearen Teilraum U^H. Die orthogonale Projektion f_H^* des R^n auf den zu U^H orthogonalen Teilraum U^K im R^H ist gegeben durch:

(3.5) $f_H^* = id - f_H$ d.h. für alle $\theta \in U^H$ gilt:

$$f_H^*(\theta) = 0 \iff \theta \in U^H \quad .$$

Die orthogonale Projektion f_H^* läßt sich darstellen durch eine idempotente, symmetrische nxn Matrix K, so daß das Testproblem (3.4) folgendermaßen dargestellt werden kann:

(3.6) H: $K\theta = 0$

A: $K\theta \neq 0$.

Die orthogonale Projektion f_H^* hat folgende wichtige Eigenschaften: Sie bildet zwei Elemente y, y^+ des R^n auf denselben Wert genau dann ab, wenn sie sich "nur" um ein Element aus U^H unterscheiden: $y = y^+ + \theta^*$, $\theta^* \in U^H$. In U^H sind aber gerade solche θ enthalten, von denen das Testproblem (3.4) nicht abhängen darf. Diese Eigenschaft von f_H^* wird als Maximalinvarianz von f_H^* bezuglich G bezeichnet; f_H^* heißt eine Maximalinvariante bezüglich G.

Das Testproblem (3.1) oder (3.4) hängt nur über die bezüglich G maximalinvariante Statistik $f_H^*(Y)$ von Y ab. Es ist "durch Invarianz reduziert" worden.

3.1.2 Beispiel für die Reduktion durch Invarianz

Am Beispiel des einfaktoriellen vollständigen Blockplanes mit festen Blockeffekten (vgl. Kirk, 1968, S. 131ff) soll obige abstrakte Darstellung erläutert werden. Üblicherweise lautet das lineare Modell der ANOVA für diesen Versuchsplan:

$$(3.7)\quad Y_{ij} = \mu+\alpha_j+\beta_i+\varepsilon_{ij} \qquad i=1,..,r;\ j=1,..,k;\ n=kr \quad .$$

Die ε_{ij} sind unabhängige, $NV(0,\sigma^2)$ -Zufallsvariablen d.h. normalverteilt mit Erwartungswert Null und Varianz σ^2, die α_j sind die Effekte der Stufen des betrachteten Faktors und die β_i sind die Effekte der Blöcke. Folgende übliche Reparametrisierungsbedingungen sollen zusätzlich gelten: $\beta_.=\alpha_.=0$.

Das Testproblem "Kein Effekt des Faktors" wird üblicherweise so formuliert:

$$\alpha_1 = \ldots = \alpha_k = 0 \quad .$$

In vektorieller Schreibweise stellt man - bei sog. lexikographischer Indizierung - den n-dimensionalen Beobachtungsvektor so dar:

$$Y = (Y_{11},..,Y_{1k},...,Y_{r1},..,Y_{rk})' \quad .$$

θ und ε werden analog indiziert. Damit hat man:

$$U := \{\theta\in R^{kr} \mid \theta_{ij}=\mu+\beta_i+\alpha_j;\ \mu,\beta_i,\alpha_j\in R;\ i=1,..,r;j=1,..,k\}$$

(3.8) und

$$U^H := \{\theta\in R^{kr} \mid \theta_{ij}=\mu+\beta_i;\ \mu,\beta_i\in R;\ i=1,..,r;j=1,..,k\} \quad .$$

Aus (3.8) kann man die zu f_H^* gehörige idempotente Matrix K herleiten:

$$\theta \in U^H \iff \theta_{11} = \theta_{12} = \dots = \theta_{ik}$$

$$\iff \theta_{ij} = \bar{\theta}_{i.}$$

(3.9) $$\iff \theta_{ij} - \bar{\theta}_{i.} = 0 \quad \text{für alle } i=1,..,r;\ j=1,..,k.$$

In Matrixschreibweise läßt sich (3.9) dann so darstellen:

(3.10) $$K\theta := (I_{rk} - 1/k I_r \otimes J_k)\theta = 0 \quad .$$

I_n bezeichnet die nxn-dimensionale Einheitsmatrix (Diagonalelemente 1, sonst 0). J_n die nxn-dimensionale Matrix bestehend aus Einsen und $\otimes$ das Kroneckerprodukt.
Mit den Rechenregeln aus Anhang A2.1 sieht man leicht, daß K idempotent ist.

Die Anwendung von K auf den Beobachtungsvektor Y ergibt komponentenweise:

(3.11) $$(KY)_{ij} = Y_{ij} - \bar{Y}_{i.} = Y_{ij} - \bar{Y}_{..} - (\bar{Y}_{i.} - \bar{Y}_{..}) \quad .$$

Die Transformation f_H^* bewirkt also, daß die als Kleinste-Quadratschätzer für die "störenden" Parameter μ, β_i bekannten Größen $\hat{\mu} = \bar{Y}_{..}$ und $\hat{\beta}_i = \bar{Y}_{i.} - \bar{Y}_{..}$,i=1,..,r, subtrahiert werden (Moosbrugger & Steyer, 1983, Kap. 4).
Bekanntlich hängt die übliche ANOVA-Teststatistik auch nur von der Maximalinvarianten f_H^* ab (Witting & Noelle, 1970). Sie lautet:

$$F(Y) = \frac{1/(k-1) \sum_{j=1}^{k} r \cdot (\bar{Y}_{.j} - \bar{Y}_{..})^2}{1/(r-1)(k-1) \sum_{i=1}^{r} \sum_{j=1}^{k} (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}$$

Wegen (3.11) hat man

$$(\bar{K}\bar{Y})_{.j} = \bar{Y}_{.j} - \bar{Y}_{..} \quad \text{und} \quad (\bar{K}\bar{Y})_{ij} - (\bar{K}\bar{Y})_{.j} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

und damit

(3.12) $$F(KY) = \frac{1/(k-1) \sum_{j=1}^{k} (\bar{K}\bar{Y})^2_{.j}}{1/(r-1)(k-1) \sum_{i=1}^{r} \sum_{j=1}^{k} ((KY)_{ij} - (\bar{K}\bar{Y})_{.j})^2}$$

3.1.3 Ableitung eines Permutationstests für lineare Hypothesen im linearen Modell

Bei Anwendung der Transformation f_H^* auf Y aus (3.1) erhält man:

(3.13) $$f_H^*(Y) = KY = K\theta + K\varepsilon \quad .$$

Bei Gültigkeit der Nullhypothese (3.6) hängt die Verteilungsfunktion von KY nicht mehr von θ ab. Sie wird von der Verteilungsfunktion von ε induziert und ist identisch mit der Verteilung von $K\varepsilon$. Für die Verteilungsfunktion F von ε war in (3.3) eine große Klasse von nicht notwendig stetigen Verteilungsfunktionen zugelassen worden. F sollte lediglich invariant bleiben gegen bestimmte Vertauschungen seiner Komponenten. Beispiele für solche Permutationsinvarianz-Strukturen werden in Abschnitt 3.1.4 gegeben.
Aus (3.3) ist bekannt, daß ε und $\pi\varepsilon$ für alle $\pi \in Q_n^*$ die gleiche Verteilung besitzen. Dasselbe gilt für $K\varepsilon$ und $K\pi\varepsilon$.
Man betrachtet nun folgende Menge von Permutationen:

$$Q_n := \{\pi^+ \in S_n \mid \text{Es gibt ein } \pi \in Q_n^* \text{ mit } \pi^+ K = K\pi\} \quad .$$

Anschaulich bedeutet die obige Eigenschaft, daß die durch π^+ bewirkte Zeilenvertauschung von K einer Spaltenvertauschung π von K entsprechen muß. Mann kann leicht zeigen, daß Q_n wieder eine Permutationsgruppe ist (Pyhel, 1978a, Kap. 1). Für Projektionen f gilt sogar mit $S_n(f) := \{\pi \in S_n \mid \pi f = f\pi\}$:

(3.14) mit
$$Q_n = S_n(f, Q_n^*) := S_n(f) \cap Q_n^*$$
$$S_n(f, Q_n^*) = \{\pi \in S_n \mid \pi f = f\rho,\ \rho \in Q_n^*\} \quad .$$

Für Permutationen $\pi^+ \in Q_n$ gilt nun nach (3.14):
Es gibt stets ein $\pi \in Q_n^*$ mit

(3.15) $$\pi^+ K\varepsilon = K\pi\varepsilon \quad .$$

Nach (3.3) haben $K\varepsilon$, $K\pi\varepsilon$ und $\pi^+K\varepsilon$ alle dieselbe Verteilungsfunktion. Somit gilt für die Verteilungsfunktion G von $K\varepsilon$:

$$G(x) = G(\pi x) \quad \text{für alle } \pi \in Q_n \text{ und alle } x \in R^n \quad .$$

Damit sind alle erforderlichen Voraussetzungen zur Konstruktion eines Permutationstests im Sinne von Kapitel 2 erfüllt. Die Permutationen $\pi^+ \in Q_n$ müssen für jedes Problem allerdings neu ermittelt werden. Eine 'geeignete' Teststatistik für das reduzierte Testproblem (3.6) wählt man auf folgende Art:

Sei U^A der bezüglich des linearen Unterraumes U aus (3.2) orthogonal komplementäre Raum zu U^H . Damit ist U^A in $U \ominus U^H$ enthalten. Jedes $\theta \in U$ ist eindeutig zerlegbar in die direkte Summe $\theta = \theta^H + \theta^A$.

θ^H enthält die 'störenden' Parameter von θ , die wegen $f_H^*(\theta^H)=0$ durch f_H^* eliminiert werden. Der Anteil $f_H^*(\theta^A)$ stellt also die Abweichung von der Nullhypothese $H:\theta \in U^H$ dar. Es gilt:

$$f_H^*(\theta^A) = 0 \iff \theta \in U^H \quad .$$

Deshalb wählt man naheliegenderweise die Projektion $f_A: R^n \to U^A$ von Y, um die unbekannte Abweichung θ^A von der Hypothese zu erfassen. Die vorher ausgeführte Transformation $f_H^*(Y)$ hat auf diese Abweichung keinen Einfluß, denn es gilt wegen $U^A \subset U^K$:

$$f_A(Y) = f_A(f_H^*(Y)) \quad . \tag{3.16}$$

Als Teststatistik eignet sich das Quadrat der (euklidischen) Länge der Projektion $f_A(Y)$:

$$S(Y) = |f_A(f_H^*(Y))|^2 \quad . \tag{3.17}$$

S(Y) ist oft identisch mit der Summe der quadrierten Abweichungen im Zähler der F-Statistik für das jeweils entsprechende ANOVA-Testproblem. 'Große' Abweichungen $f_A(\theta)=\theta^A$

von 0 führen zu 'großen' Werten der Teststatistik S. Auch eine quadratische Form in $f_A(Y)$ als Teststatistik ist möglich (vgl. Pyhel, 1980a).

Ein Permutationstest wird nun, wie in Kapitel 2 beschrieben, durchgeführt. Man ermittelt die Invarianzeigenschaften von $f_H^*(Y)$, d.h. man bestimmt die Permutationsgruppe Q_n aus (3.14). Dann wird der transformierte Beobachtungsvektor $t=f_H^*(y)$ gebildet und $S(t)=|f_A(t)|^2$ berechnet. Anschließend werden für alle Permutationen $\pi\in Q_n$ die Werte $S(\pi t) = |f_A(\pi t)|^2$ ermittelt und die Größen $p^>$ und $p^=$ bestimmt:

$$(3.18) \qquad p^> = |\{\pi \in Q_n |\ S(\pi t) > S(t)\}|/|Q_n|$$
$$p^= = |\{\pi \in Q_n |\ S(\pi t) = S(t)\}|/|Q_n| \quad .$$

Der nicht randomisierte Test zum Niveau $\leq\alpha$, verwirft die Nullhypothese H, falls der p-Wert $p=p^>+p^=\leq\alpha$ ist.

3.1.4 Ein Invarianz-Permutationstest für den einfaktoriellen Blockplan

Für den Fehlervektor $\varepsilon=(\varepsilon_{11},\dots,\varepsilon_{1k},\dots,\varepsilon_{r1},\dots,\varepsilon_{rk})'$ wird in der ANOVA mit festen Effekten angenommen, daß alle Komponenten ε_{ij} unabhängig und identisch verteilt sind nach einer $NV(0,\sigma^2)$-Verteilung.

Auch unter Vernachlässigung der Normalverteilungsannahme bedeutet das für die Verteilungsfunktion $F(x)$, $x\in R^n$ von ε:

(3.19) $F(x) = F(\pi x)$ für alle $(rk)!$ Permutationen π der symmetrischen Gruppe S_{rk} der Ordnung rk .

Falls die Fehlerkomponenten $\varepsilon_i=(\varepsilon_{i1},..,\varepsilon_{ik})', i=1,..,r$, die alle zu demselben Block gehören, stochastisch abhängig sind, die verschiedenen ε_i aber unabhängig und identisch verteilt, kann F nicht invariant sein unter Permutationen, die zwei Komponenten ε_{ij} und $\varepsilon_{i'j'}$ aus verschiedenen Blöcken ($i\neq i'$) miteinander vertauschen. Durch solch eine Permutation würde die Struktur der Abhängigkeiten innerhalb der Blöcke i und i' verändert. Die stochastische Abhängigkeit je Block soll aber die Permutationsinvarianz pro Block nicht beeinflussen. Dann bleiben pro Block k! Permutationen und damit insgesamt die Untergruppe Q^*_{rk} von S_{rk} mit $(k!)^r \cdot r!$ Permutationen. Der Faktor r! beruht darauf, daß wegen der identischen Verteilung der ε_i alle r! Vertauschungen ganzer Blockvektoren ε_i zulässig sind.

Q^*_{rk} läßt sich darstellen als das sog. Kranzprodukt von S_k mit S_r :

(3.20) $$Q^*_{rk} = S_k \wr S_r$$

(vgl. Bell & Haller, 1969; Bell & Smith, 1969, Kerber, 1971).

Gibt man schließlich die Annahme gleicher Verteilungen für die ε_i, i=1,..,r,auf, bleibt als Permutationsgruppe Q^{**}_{rk} mit Vertauschungen ausschließlich innerhalb der Blöcke. Q^{**}_{rk} hat die Ordnung $(k!)^r$ und wird als direktes Produkt von r Kopien von S_k bezeichnet:

(3.20b) $$Q^{**}_{rk} = \bigtimes_{i=1}^{r} S_{k,i}$$

(vgl. Haller, 1969; Kerber, 1971, S. 25).

Nach (3.14) muß $\{\pi \in S_{rk} \mid \pi K = K\pi\}$ bestimmt werden mit der orthogonalen Projektionsmatrix K (vgl. (3.10)):

$$K = I_{rk} - 1/k I_r \otimes J_k = I_r \otimes (I_k - 1/k J_k) \quad .$$

Aus der Blockdiagonalstruktur von K ist ersichtlich, daß Zeilenvertauschungen zwischen verschiedenen Blocken nicht durch Spaltenvertauschungen kompensierbar sind. Zeilenweise Vertauschungen innerhalb der Blocke dagegen sind durch identische spaltenweise Vertauschungen herstellbar. Zeilenweise Vertauschungen ganzer Diagonalblocke sind trivialerweise durch spaltenweise Vertauschungen derselben Blocke auszugleichen, da die Matrix Blockdiagonalgestalt hat. Die Menge aller dieser $(k!)^r \cdot r!$ Elemente ist also ebenfalls $S_k \wr S_r$. Q_n ist also wegen (3.14) ebenfalls identisch mit $S_k \wr S_r$ und repräsentiert die Invarianzstruktur von K_ε.

Abschließend ist noch f_A zu ermitteln. In $\theta_{ij}=\mu+\beta_i+\alpha_j$ ist α_j der nicht in U^H enthaltene Anteil.

Es gilt mit der Reparametrisierungsbedingung der Modellgleichung (3.7):

$$\bar{\theta}_{.j} - \bar{\theta}_{..} = \alpha_j \quad .$$

In Matrixschreibweise hat man damit für f_A:

$$f_A(\theta) = K_A\theta = 1/rJ_r \otimes (I_k - 1/kJ_k) \quad .$$

Somit ergibt sich

$$K_A(KY) = ((\bar{K}\bar{Y})_{.1},..,(\bar{K}\bar{Y})_{.k},....,(\bar{K}\bar{Y})_{.1},..,(\bar{K}\bar{Y})_{.k})' \quad ,$$

denn $(\bar{K}\bar{Y})_{..}=0$ wegen (3.11).

Damit hat die Teststatistik folgende Gestalt:

$$(3.21) \quad S(Y) = |f_{A_k}(KY)|^2 = r \sum_{j=1}^{k} (\bar{K}\bar{Y})^2_{.j} = r \sum_{j=1}^{k} (\bar{Y}_{.j} - \bar{Y}_{..})^2$$
$$= r\left(\sum_{j=1}^{k} \bar{Y}^2_{.j} - k\bar{Y}^2_{..} \right) \quad .$$

Wie schon erwähnt, ist S identisch mit der Summe der quadrierten Abweichungen im Zähler der entsprechenden ANOVA-Teststatistik.

Für die "praktische" Durchführung des Permutationstests überlegt man sich leicht, daß die r! Vertauschungen ganzer Blöcke die $\bar{Y}_{.j}$ nicht verändern und $\bar{Y}_{..}$ unter allen Permutationen aus Q_n (und natürlich auch S_n) gleich bleibt. Es

bleiben somit $(k!)^r$ Intrablockpermutationen übrig (vgl. Hoeffding, 1952; Scheffe, 1959).

3.1.5 Invarianz-Permutationstests für verschiedene mehrfaktorielle Versuchspläne

Das zuvor vorgestellte Verfahren bietet einen allgemeinen Lösungsansatz zur Konstruktion von Invarianz-Permutationstests. Wie in Abschnitt 3.1.3 erläutert, ist die Permutationsgruppe Q_n, die die Invarianzeigenschaften des transformierten Fehlervektors charakterisiert, abhängig von dieser Transformation f_H^* bzw. der zugehorigen Matrix K. Es ist damit für jedes Testproblem in einem bestimmten linearen Modell neu zu prüfen, ob Q_n "hinreichend viele" Permutationen zur Konstruktion eines (nicht-trivialen) Permutationstests etwa mit der Teststatistik $S=|f_A(f_H^*(Y))|$ enthält.

Für den Test auf Wechselwirkungen im zweifaktoriellen, vollständigen Zufallsplan (CRF - cr in der Notation von Kirk, 1968) z.B. existiert kein solcher Invarianz-Permutationstest. Dieses Ergebnis ist in Pyhel (1978a) und Willmes (1981b) nur kurz erwähnt; auch Edgington (1980, S. 181f) bietet nur eine 'heuristische' Erklärung. In Kendall & Stuart (1968, S. 108) wird ebenfalls vermutet, daß ein Wechselwirkungs-Test nicht existiert.

Zur Vereinfachung der Darstellung wird gleiche Zellbesetzung $n>1$ angenommen. Die Modellgleichung lautet:

$$Y_{ijk} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\varepsilon_{ijk} \qquad \begin{matrix} i=1,..,r;\ j=1,..,c; \\ k=1,..,n;\ N=rcn \end{matrix} \tag{3.22}$$

mit den üblichen Reparametrisierungsbedingungen:

$$\alpha_{\cdot}=\beta_{\cdot}=(\alpha\beta)_{i\cdot}=(\alpha\beta)_{\cdot j}=(\alpha\beta)_{\cdot\cdot}=0$$

und der 'schwachen' Verteilungsannahme für die Verteilungsfunktion F des Versuchsfehlers $\varepsilon=(\varepsilon_{111},\ldots,\varepsilon_{11n},\ldots,\varepsilon_{rcn})'$:

(3.23) $\quad F(x) = F(\pi x)$ für alle $\pi\in S_N$ und alle $x\in R^N$.

In Matrixschreibweise lautet die Modellgleichung (3.22):

$$Y = Xb + \varepsilon \quad \text{oder} \quad Y = \theta + \varepsilon\ ,\ \theta\in U$$

$$\text{mit } U=\{\theta\in R^n \mid \theta=Xb, b\in R^{rc}\}\ .$$

Der Parametervektor b hat folgende Gestalt:

$$b=(\mu,\alpha_1,\ldots,\alpha_{r-1},\beta_1,\ldots,\beta_{c-1},(\alpha\beta)_{11},\ldots,(\alpha\beta)_{1(c-1)},\ldots,(\alpha\beta)_{(r-1)(c-1)})'\ .$$

Den Versuchsplan kann man sich z.B. wie in Kirk (1968, Kap. 7) veranschaulichen.

Die Design-Matrix X hat vollen (Spalten-) Rang rc.

(3.24) $\quad X = (1_N \mid M_r\otimes 1_{cn} \mid 1_r\otimes M_c\otimes 1_n \mid M_r\otimes M_c\otimes 1_n)$

(Zur Definition von M_r siehe Anhang A2)

Das Design des Versuchsplans (vgl. Kendall & Stuart, 1968, S. 514f) ist orthogonal, d.h. X läßt sich in m (= 4) Teile zerlegen:

(3.25) $\quad X = (\overset{\circ}{X}_1 \mid \overset{\circ}{X}_2 \mid \ldots \mid \overset{\circ}{X}_m)$ mit

$$\overset{\circ}{X}_l'\overset{\circ}{X}_{l'} = 0 \quad \text{für alle } l\neq l'\in\{1,\ldots,m\}\ .$$

Mit der Definition

$$X_l := (0\mid\ldots\mid\overset{\circ}{X}_l\mid 0) \qquad l=1,\ldots,m$$

gilt:

$$X = \sum_{l=1}^{m} X_l \quad \text{und } X_l'X_{l'}=0 \text{ für alle } l\neq l'\ .$$

Die Unterräume des R^n

$$U := \{\theta_l \mid \theta_l = X_l b,\ b\in R^{rc}\}$$

sind dann ebenfalls paarweise orthogonal und U ist die direkte Summe dieser Unterräume (vgl. Anhang A4).

$$(3.26) \qquad U = \bigoplus_{l=1}^{m} U_l \quad .$$

Das Testproblem lautet dann:

$$(3.27) \qquad H: \quad X_{l_o} b = \theta_{l_o} = 0 \quad \text{für ein } l_o \in \{1,..,m\}$$

$$A: \quad \theta_{l_o} \neq 0 \quad .$$

Für das zu betrachtende Testproblem ist $l_o=4$ mit X_4 aus (3.24).

Mit Hilfe der Teilräume läßt sich äquivalent formulieren:

$$(3.28) \qquad H: \quad Xb = \theta \in U^H \text{ mit } U^H := \bigoplus_{l \neq l_o}^{m} U_l$$

$$A: \quad \theta \in U \setminus U^H \quad .$$

Wie zuvor in Abschnitt 3.1.1 erläutert, soll der Test für das Problem (3.27) nicht von $\sum_{l \neq l_o}^{m} \theta_l$ abhängen:
Das Testproblem ist invariant gegen

$$G_{l_o} = \{g: R^N \to R^N \mid g = id + \theta, \ \theta \in U^H = \bigoplus_{l \neq l_o}^{m} U_l\} \quad .$$

Die bezüglich G_{l_o} maximalinvariante orthogonale Projektion f^*_{H,l_o} des R^N auf den zu U^H orthogonalen Raum in R^N ist gegeben durch:

$$(3.29) \qquad f^*_{H,l_o} = id - \sum_{l \neq l_o} f_l$$

mit

$$f_l: R^N \to U_l \qquad \text{Projektion des } R^N \text{ auf } U_l.$$

(vgl. Pyhel, 1978a).

Für die Summe der f_l gilt: $f_l: R^N \to U$ ist eine Projektion des R^N auf $U = \bigoplus_{l=1}^{m} U_l$.
Sie ist gegeben durch $X(X'X)^{-1}X'$ (Moosbrugger & Steyer, 1983).

Entsprechend sind die einzelnen Projektionen $f_l: R^N \to U_l$ gegeben durch: $X_l(X'X)^{-1}X'$,

und damit ist f^*_{H,l_o} darstellbar durch die Matrix

$$K_{l_o} = I_N - (X - X_{l_o})(X'X)^{-1}X' \quad .$$

Es gilt wiederum

$$K_{l_o} Y = X_{l_o} b + K_{l_o} \varepsilon \qquad \text{und}$$

$$K_{l_o} Y = K_{l_o} \varepsilon \qquad \text{unter H: } X_{l_o} b = 0 \quad .$$

In der Teststatistik

$$S(Y) = |f_A(f^*_{H,l_o}(Y))|^2 = |f_A(Y)|^2$$

wird $f_A: R^N \to U^A$, die Projektion in den bezuglich U orthogonalen Unterraum von U^H benötigt.

Wegen (3.26) ist $U^A = U_{l_o}$, und die zu f_A zugehörige Projektionsmatrix ist $X_{l_o}(X'X)^{-1}X'$.

Für die Teststatistiken der ANOVA mit orthogonalem Design und festen Effekten gilt:

$$F(Y) = \frac{|f_{H,l_o}(Y)|^2 / \dim U_{l_o}}{|Y - \sum_{l=1}^{m} f_l(Y)|^2 / (N - \dim U)} \qquad \text{und } F(Y) = F(f^*_{H,l_o}(Y)) \; .$$

Also ist die F-Statistik auch eine Funktion der Maximalinvarianten f^*_{H,l_o} .

Für den Test auf Wechselwirkungen in CRF-rc-Versuchsplan bedeuten obige Ergebnisse:

$$\theta_{l_o} = \theta_4 = (0|0|0|M_r \otimes M_c \otimes 1_n) b$$

$$U = \{\theta \in R^N \mid \theta_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \; ; \; \mu, \alpha_i, \beta_j, (\alpha\beta)_{ij} \in R; \; i=1,..,r; j=1,$$

$$U^H = \{\theta \in U \mid \theta_{ijk} = \mu + \alpha_i + \beta_j\}$$

und

$$K_4 = I_N - (1_N | M_r \otimes 1_{cn} | 1_r \otimes M_c \otimes 1_n | 0)(X'X)^{-1}X'$$

$$= I_N - (I_r \otimes 1/cnJ_{cn} + 1/rJ_r \otimes I_c \otimes 1/nJ_n - 1/NJ_N)$$

$$(3.30) = I_N - 1/cnI_r \otimes J_{cn} - 1/rnJ_r \otimes I_c \otimes J_n + 1/NJ_N \quad .$$

Für die Komponenten von $K_4 Y$ gilt:

$$(K_4 Y)_{ijk} = T_{ijk} = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} = 0$$

falls

$$\theta_{ijk} = (X_4 b)_{ijk} = \theta_{ijk} - \bar{\theta}_{i..} - \bar{\theta}_{.j.} + \bar{\theta}_{...} = 0 \quad .$$

Zur Bestimmung der Invarianzstruktur von K_4^{ε} muß $\mathcal{Q}_N$ nach (3.14) ermittelt werden. Nach (3.23) ist $\mathcal{Q}^*_N = S_N$ und damit

$$\mathcal{Q}_N = S_N(K_4) \quad .$$

K_4 läßt sich auch schreiben als:

(3.31) $$I_N - (1/cI_r \otimes J_c + 1/rJ_r \otimes I_c) \otimes 1/nJ_n + 1/NJ_N \quad .$$

Mit den von Pyhel (1978a, S. 11) und Stein-Zaar (1983, S. 15) angegebenen Regeln findet man:

(3.32) $$\mathcal{Q}_N = S_n \wr (S_r \otimes S_c) \quad .$$

Dabei ist

$$S_r \otimes S_c := \{\pi \otimes \rho \mid \pi \in S_r, \rho \in S_c\}$$

und $\pi \otimes \rho$ bezeichne das Kronecker-Produkt der beiden Matrizen, die π und ρ darstellen. Die Ordnung von $S_r \otimes S_c$ ist $r!c!$.

In $\mathcal{Q}_N$ sind solche Permutationen enthalten, die für

$$(T_{111},\dots,T_{11n},\dots,T_{121},\dots,T_{12n},\dots,T_{rc1},\dots,T_{rcn})'$$

Vertauschungen von Elementen innerhalb der 'Zellen', von ganzen Zeilen und von ganzen Spalten des Versuchsplans bewirken.

Betrachtet man die Teststatistik

$$\begin{aligned} S(Y) &= |f_A(K_4 Y)|^2 = |f_A(Y)|^2 = |X_4 (X'X)^{-1} X'Y|^2 \\ &= \sum_{i=1}^{r} \sum_{j=1}^{c} n(1/n \sum_{k=1}^{n} T_{ijk})^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} n\bar{T}^2_{ij.} \\ &= \sum_{i=1}^{r} \sum_{j=1}^{c} n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \end{aligned} \quad (3.33)$$

so erkennt man sofort, daß S(Y) invariant ist gegenuber allen $(n!)^{rc} r!c!$ Permutationen aus $\mathcal{Q}_N$ in (3.32).

Damit ist mit diesen Methoden kein nicht-trivialer Invarianz-Permutationstest für die Wechselwirkungshypothese im zweifaktoriellen Zufallsplan angebbar. Die in der Literatur berichteten Tests in Randomisierungsmodellen von Wilk & Kempthorne (1956) werden in Abschnitt 3.3 diskutiert.

Ohne die obige etwas aufwendige Begründung kommt man durch 'heuristische' Überlegungen wie in Pyhel (1980b) zu demselben Ergebnis. In dem linearen Modell mit (3.2) und (3.3) soll das Testproblem

(3.34) $\quad H: K\theta = 0 \quad$ gegen $A: K\theta \in U^A \subseteq R^N \setminus \{0\}$

getestet werden. Mit dem Residuen-Vektor $\varepsilon := Y - \theta$ betrachtet man die Transformation $T: R^N \to R^N$ (vgl. (3.13)) von Y

(3.35) $\quad T(Y) := KY = K\theta + K\varepsilon \quad .$

Zuvor war diese Transformation durch bestimmte Maximalinvarianz-Forderungen gerechtfertigt worden.

Im Beispiel des CRF-rc nach (3.22) war:

$$U = \{\theta = \Delta \otimes 1_n \mid \Delta \in R^{cr},\ \Delta_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij};\ \mu, \alpha_i, \beta_j, (\alpha\beta)_{ij} \in R\}$$

und

$$U^H = \{\theta = \Delta \otimes 1_n \mid \Delta \in R^{cr},\ \Delta_{ij} = \mu + \alpha_i + \beta_j;\ \mu, \alpha_i, \beta_j \in R\} \quad .$$

Nun gilt wegen der üblichen Reparametrisierungsbedingungen:

$$\theta \in U^H \iff \Delta_{ij} \otimes 1_n = (\bar{\Delta}_{i.} + \bar{\Delta}_{.j} - \bar{\Delta}_{..}) \otimes 1_n$$

$$\iff (\Delta_{ij} - \bar{\Delta}_{i.} - \bar{\Delta}_{.j} + \bar{\Delta}_{..}) \otimes 1_n = 0 \qquad \text{für alle}$$

$$i = 1,..,r;\ j = 1,..,c \ .$$

Aus der Struktur der obigen Gleichung leitet man direkt die obige Transformationsmatrix K ab (vgl. (3.11)):

$$K = I_{rcn} - (1/cI_r \otimes J_c - 1/rJ_r \otimes I_c + 1/rcJ_r \otimes J_c) \otimes 1/nJ_n$$

mit

$$(KY)_{ijk} := T_{ijk} = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} \quad .$$

Jede 'sinnvolle' Teststatistik basiert aber auf den 'Zellen'-Mittelwerten $\bar{T}_{ij.}$ (vgl. (3.33)). Diese ist jedoch invariant gegenüber den zulässigen Permutationen (3.32). Der von Still & White (1981) vorgeschlagene Test läßt sich damit nicht als Invarianz-Permutationstest begründen.

Dieses negative Ergebnis gilt aber nicht für alle mehrfaktoriellen Pläne. Im Folgenden werden mehrere wichtige Versuchspläne angeführt, für die Invarianz-Permutationstests für (die) Haupteffekte oder die (hochsten) Wechselwirkungen existieren.

(i) Unverbundenes Zweistichprobenproblem (Pitman, 1937a; Lehmann & Stein, 1949; Scheffe, 1959; Witting & Nölle, 1970; Conover, 1971, Maritz, 1981; Willmes & Pyhel, 1981):

Das Experiment wird beschrieben durch

$Y = (Y_{11},..,Y_{1n_1},Y_{21},.....,Y_{2n_2})'$ mit Y_{ij} jeweils identisch verteilt nach F_i , $i=1,2$; $j=1,..,n_i$ und $F_1(x) = F_2(x-\delta)$ mit $\delta,x\in R$.

Die Nullhypothese ist H: $\delta = 0$ gegen A: $\delta \neq 0$.

Eine Teststatistik für den zweiseitigen Test ist z.B.:

(3.36) $$S(Y) = (\bar{Y}_{1.}-\bar{Y}_{2.})^2 \quad .$$

Unter der Nullhypothese ist Y invariant gegen alle $(n_1+n_2)!$ Vertauschungen seiner Komponenten. Nur $(n_1+n_2)!/n_1!n_2!$ dieser Permutationen können zu unterschiedlichen Werten der Teststatistik S führen. Die (nicht randomisierte) Testentscheidung wird wie im Anschluß an (3.18) erläutert getroffen.

Man hätte auch die übliche Teststatistik des t-Tests wählen können. S und t stehen in monotoner Beziehung zueinander (Lehmann, 1959, S. 188f; Kendall & Stuart, 1981):

(3.37) $$(\bar{Y}_{1.}-\bar{Y}_{2.})^2 = n^2s^2/(n_1n_2(1+(n-2)/t^2))) \quad \text{mit}$$

$$s^2 = 1/(n-1)\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{..})^2 \quad \text{und} \quad n=n_1+n_2 \quad .$$

Damit reicht es aus, (3.36) zu betrachten.

Bei einseitiger Fragestellung: H: $\delta\leq 0$ gegen A: $\delta>0$ genügt es,

(3.38) $S(Y) = \sum_{j=1}^{n_1} Y_{1j}$ zu betrachten.

Man kann auch die Teststatistik (3.38) für den zweiseitigen Test nutzen, wenn die Teststatistik einmal für -y und dann für +y mit allen zulässigen Permutationen $S(-\pi y)$ bzw. $S(\pi y)$ verglichen wird. H ist zu verwerfen, falls

$$p = 2 \cdot \min\{(p^{>}+p^{=})_{+},\ (p^{>}+p^{=})_{-},\ 1/2\} \leq \alpha \quad \text{ist.}$$

(ii) Verbundenes Zweistichprobenproblem Lehmann, 1959, 1975; Witting & Nölle, 1970; Conover, 1971; Maritz, 1981; Willmes & Pyhel, 1981)

Das Experiment wird beschrieben durch

$Y = (Y_{11}, Y_{21}, \ldots, Y_{1n}, Y_{2n})'$.

Sei $Y_i := Y_{1i} - Y_{2i}$ verteilt nach symmetrischem $F_i(x-\delta)$, i=1,..,n. Es wird also der Spezialfall einer Verschiebealternative (sog. 'shift-model'; vgl. Lehmann, 1975, S. 156ff) betrachtet. Auf die kritischen Einwände von Krauth (1982) sei hingewiesen.

Unter der Nullhypothese ist die Verteilung von Y invariant gegen alle 2^n Vertauschungen der Komponenten in jedem der n 'Blöcke' mit je 2 Beobachtungen. Geeignete Teststatistiken für den zweiseitigen bzw. einseitigen Test sind:

(3.39) $S(Y) = \left| \sum_{i=1}^{n} (Y_{1i} - Y_{2i}) \right|$ bzw. $S(Y) = \sum_{i=1}^{n} Y_{1i}$.

Auch hier besteht wieder ein monotoner Zusammenhang zur üblichen Statistik des t-Tests.

Sen (1965) hat einen Permutationstest mit verallgemeinerten U-Statistiken angegeben.

(iii) Unverbundenes k-Stichprobenproblem (Einfaktorieller Zufallsplan, CR-k nach Kirk, 1968; Box & Andersen, 1955, Sen, 1965, 1967; Silvey, 1970; Maritz, 1981; Willmes & Pyhel, 1981)

Das Experiment wird beschrieben durch:

$$Y = (Y_{11},..,Y_{1n_1},Y_{21},..,Y_{2n_2},....,Y_{k1},..,Y_{kn_k})'$$

mit Y_{ij} jeweils für $j=1,..,n_j$ verteilt nach

$$F_i(x) = F(x-\mu_i) \quad \text{für alle } i=1,..,k;\ n=\sum_{i=1}^{k} n_i \ .$$

Die Nullhypothese lautet:

$$H: \mu_i = \mu \quad \text{gegen} \quad A: \mu_i \neq \mu_{i'} \quad \text{für mindestens ein } i\neq i' \in \{1,..,k\} \ .$$

Eine geeignete Teststatistik ist:

$$(3.40) \quad S(Y) = \sum_{i=1}^{k} n_i \bar{Y}^2_{i.} = \sum_{i=1}^{k} Y^2_{i.}/n_i \quad \text{bzw.} \quad S(Y)= \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \ .$$

Unter der Nullhypothese ist Y invariant gegen alle n! Vertauschungen seiner Komponenten. Nur $n!/\prod_{i=1}^{k} n_i!$ von ihnen können zu unterschiedlichen Werten der Teststatistik führen. $\bar{Y}_{..}$ ist invariant unter allen zulässigen Permutationen.
Sen (1967) gibt eine ganze Klasse von Teststatistiken an, die auf verallgemeinerten U-Statistiken beruhen; (3.40) ist davon ein Spezialfall (für r=0 in Formel (5.5), S. 1208).

(iv) Verbundenes k-Stichprobenproblem (einfaktorieller Blockplan; RB-k nach Kirk, 1968, Welch, 1937, Pitman, 1938; Wilks, 1962; Maritz, 1981; Willmes & Pyhel, 1981)

Ein Permutationstest für diesen Versuchsplan wurde bereits im Abschnitt 3.1.4 hergeleitet mit der Teststatistik

(3.41) $S(Y) = \sum_{j=1}^{k} Y^2_{\cdot j}$.

Unter der Nullhypothese 'gleicher Effekt der k Faktorstufen' sind alle $(k!)^r$ Intrablockpermutationen gleichwahrscheinlich.

(v) Verbundenes k-Stichprobenprolem mit Wiederholungen

Falls pro Faktorstufe in jedem Block mehr als eine Beobachtung vorliegt, hat man das Modell:

(3.42) $Y_{ijl} = \mu+\beta_i+\alpha_l+\varepsilon_{ijl}$ $\quad i=1,..,r$ Blöcke, $j=1,..,n_{il}>1$ Replikationen, $l=1,..,k$ Faktorstufen mit $\beta_\cdot=\alpha_\cdot=0$ und $n=\sum_i\sum_l n_{il}$

- in Erweiterung von Abschnitt 3.4.1 - mit folgender Vertauschbarkeitsannahme für die Fehlerkomponenten je Block:

$$\varepsilon_i=(\varepsilon_{i11},\dots,\varepsilon_{in_{i1}1},\dots,\varepsilon_{i1k},\dots,\varepsilon_{in_{ik}k})' \sim F_i(x),\ x\in R^{n_i\cdot}$$

mit $F_i(x)=F_i(\pi x)$ für $\pi\in S_{n_i\cdot}$.

Insgesamt hat man die Permutationsgruppe der $\prod_{i=1}^{r} n_i!$ Intrablockpermutationen.

Im Unterschied zu Wilk (1955), Cox & Hinckley (1974, S. 205) oder Addelman (1970) werden keine festen Faktor x Block-Interaktionen in die Modellgleichung aufgenommen. Das Modell (3.42) entspricht dem Modell in Krafft (1978, S. 179, 195 mit $n_{il} > 1$).

In Verallgemeinerung von (3.9) und (3.10) hat man als Transformationsmatrix:

(3.43) $K = I_n - \bigoplus_{i=1}^{r} 1/n_{i\cdot} J_{n_{i\cdot}} = \bigoplus_{i=1}^{r} (I_{n_{i\cdot}} - 1/n_{i\cdot} J_{n_{i\cdot}})$.

K ist wieder idempotent, da die $I_{n_{i\cdot}} - 1/n_{i\cdot} J_{n_{i\cdot}}$ es sind.

Analog zum Fall ohne Replikationen gilt:

$$K_A = \bigoplus_{l=1}^{k} 1/n_{.l} J_{n_{.l}} - 1/nJ_n \quad \text{und}$$

$$S(Y) = |K_A(KY)|^2 = \sum_{l=1}^{k} n_{.l}(\bar{Y}_{..l} - \bar{Y}_{...})^2 \quad \text{oder}$$

äquivalent, da $\bar{Y}_{...}$ invariant ist für alle unter der Nullhypothese zulässigen $\prod_{i=1}^{r} n_{i.}!$ Intrablockpermutationen:

$$S(Y) = \sum_{l=1}^{k} n_{.l} \bar{Y}^2_{..l} \quad . \tag{3.44}$$

(vi) Ausgewogener vollständiger Blockplan (BIBD; Mitra, 1960)

In einem vollständigen Blockplan werden in jedem Block alle Stufen des interessierenden (Behandlungs-) Faktors (mindestens) einmal auf eine Beobachtungseinheit angewendet. Es ist aber auch möglich, daß nicht alle Faktorstufen für einen Block vorkommen sollen bzw. können. In einem ausgewogenen unvollständigen Blockplan (Kirk, 1968, Kap. 11.1 - 11.4; Winer, 1971; Scheffe, 1959, S. 161f) sind alle möglichen Paare von Faktorstufen über alle r Blöcke der Länge c hinweg gleich oft (λ - mal) vertreten. Bei insgesamt k Faktorstufen müssen dann folgende Bedingungen erfüllt sein:

$k(k-1)\lambda = rc(c-1)$ und $k \leq r$. Jede Faktorstufe kommt insgesamt in s Blöcken vor mit $ks = rc$. In Cochran & Cox (1957, Kap. 11) oder Fisher & Yates (1963) sind für praktisch wichtige Kombinationen von λ, k, r, c, ausgewogene unvollständige Blockpläne angegeben. Falls $km=r$; $m \geq 1$, läßt sich auch stets ein BIBD angeben, in dem jede Faktorstufe gleich häufig in jeder der c Positionen je Block vorkommt (vgl. Scheffe, 1959, S. 169).

Das Modell lautet dann:

(3.45) $Y_{ij} = \mu+\beta_i+\alpha_j+\varepsilon_{ij}$ mit $(i,j)\in D$; $D:=\{L_{ij}|L_{ij}=1\}$

wobei $L_{ij}=\begin{cases}1 & \text{Faktorstufe } j \text{ tritt in Block } i \text{ auf}\\ 0 & \text{sonst}\end{cases}$

und den Reparametrisierungsbedingungen $\beta_\cdot=\alpha_\cdot=0$.

Die Vertauschbarkeitsannahme für die c-dimensionalen $(c<k)$ Fehlerkomponenten $\varepsilon_i=(\varepsilon_{i1},..,\varepsilon_{ij},..,\varepsilon_{ik})'$, $(i,j)\in D$ je Block ist:

$$F_i(x) = F_i(\pi x) \quad \text{mit} \quad \pi\in S_{L_i};\ L_i=(L_{i1},..,L_{ik})' \text{ und } |S_{L_i}|=c!$$

Insgesamt hat man $(c!)^r$ Intrablockpermutationen in $\times_{i=1}^{r} S_{L_i}$.

Es werden also pro Block immer nur die c Komponenten vertauscht, für die $L_{ij}=1$ ist. Analog zu Abschnitt 3.1.2 hat man:

$$U = \{\theta\in R^{rk} \mid \theta_{ij}=\mu+\beta_i+\alpha_j;\ \mu,\beta_i,\alpha_j\in R;\ i=1,..,r;j=1,..,k;(i,j)\in D\}$$

$$U^H = \{\theta\in U \mid \theta_{ij}=\mu+\beta_i;\ (i,j)\in D\} \quad .$$

Die zu f^*_H gehörige idempotente Matrix K leitet man wieder so ab:

$$\begin{aligned}
\theta\in U^H &\Leftrightarrow \theta_{i1}=\theta_{i2}=\ \dots\ =\theta_{ij}=\ ..\ =\theta_{ik} \quad \text{für alle } (i,j)\in D\\
&\Leftrightarrow \theta_{ij} = 1/c \sum_{j'=1}^{k} L_{ij'}\theta_{ij'} \quad \text{für alle } (i,j)\in D\\
&\Leftrightarrow L_{ij}(\theta_{ij}-1/c \sum_{j'} L_{ij'}\theta_{ij'}) = 0\\
&\Leftrightarrow L(I_{rk}-1/c \bigoplus_{i=1}^{r} L_iJ_k) = 0 \ ;\ L_i=\begin{pmatrix} L_{i1} & & 0\\ & \ddots & \\ 0 & & L_{ik}\end{pmatrix}_{k\times k}, \ i=1,..,r; \text{ und } L=\bigoplus_{i=1}^{r} L_i .
\end{aligned}$$

Also insgesamt:

$$K = L - 1/c \bigoplus_{i=1}^{r} L_iJ_k \quad \text{(idempotent)}$$

und $(KY)_{ij} = Y_{ij}-1/c \sum_{j'=1}^{k} L_{ij'}Y_{ij'}$ für $(i,j)\in D$.

Wie in Abschnitt 3.1.4 wird f_A aus $\bar\theta_{\cdot j}-\bar\theta_{\cdot\cdot}=\alpha_j$ für $(i,j)\in D$ ermittelt zu:

$$1/s \sum_{i=1}^{r} L_{ij}\theta_{ij} - 1/sk \sum_{i=1}^{r}\sum_{j'=1}^{k} L_{ij'}\theta_{ij'} \text{ für alle } j=1,..,k$$

$$\Leftrightarrow 1/s \sum_{i=1}^{r}[L_{ij}\theta_{ij} - 1/k \sum_{j=1}^{k} L_{ij}\theta_{ij}] \Leftrightarrow 1/s\Big(\prod_{j=1}^{k}\Longrightarrow_{i=1}^{r} L_i - 1/k\cdot LJ_{rk}\Big)\theta =: K_A\theta$$

Durch Ausrechnen findet man:

$$K_A \cdot K = \left(\prod_{j=1}^{k} \sum_{i=1}^{r} 1/s(I_k - 1/kJ_k)L_i \right)$$

und damit:

$$S(Y) = r \sum_{j=1}^{k} (\bar{Y}_{.j} - \bar{Y}_{..})^2 \quad \text{mit } \bar{Y}_{.j} = 1/s \sum_{i=1}^{r} L_{ij} Y_{ij} \text{ und}$$

$$\bar{Y}_{..} = 1/sk \sum_{i=1}^{r} \sum_{j=1}^{k} L_{ij} Y_{ij}$$

oder äquivalent

$$S(Y) = \sum_{j=1}^{k} \bar{Y}^2_{.j} \quad . \tag{3.46}$$

Unvollständige Blockpläne sind immer dann besonders geeignet, wenn die Anwendung aller Faktorstufen auf die betrachteten Versuchseinheiten zu Komplikationen führt.

In der ANOVA muß für den Test auf einen Effekt des interessierenden Faktors eine 'adjustierte' Summe quadrierter Abweichungen verwendet werden, da durch die in für verschiedene Faktorstufen fehlenden Beobachtungen in mehreren Blöcken Faktorstufen- und Blockeffekte in der Summe der Beobachtungen je Faktorstufe konfundiert sind (vgl. Kirk, 1968, Kap. 11.4). Im Permutationstest sind die 'störenden' Blockeffekte durch die Transformation eliminiert worden.

Eine Übertragung auf den Fall eines partiell ausgewogenen, unvollständigen Blockplanes (PBIBD) ist möglich (vgl. Rao, 1963; Giri, 1965; Cleroux, 1969 für Randomisierungstests).

(vii) Einfach hierarchischer zweifaktorieller Versuchsplan (CRH-r(c) nach Kirk, 1968, Kap. 7.15)

In einem hierarchischen Versuchsplan sind die Faktoren nicht mehr gekreuzt. Ein fester/zufälliger Faktor B ist unter den Stufen eines festen Faktors A 'genestet', d.h. jede Stufe des Faktors B tritt nur bei genau einer Stufe von A auf (Beispiel aus der Pädagogischen Psychologie: Faktor B: Klassen, Faktor A: Lehrmethoden). Die Modellgleichung ist:

(3.47) $Y_{ij(i)k} = \mu+\alpha_i+\beta_{j(i)}+\varepsilon_{ij(i)k}$

mit $i=1,..,r$ Stufen Faktor A; $j(i)=1,..,c(i)$ Stufen Faktor B für die i-te Stufe von A; $k=1,..,n_{ij(i)}$ Replikationen.

Weiterhin sei $n_{i.} = \sum_{j(i)=1}^{c(i)} n_{ij(i)}$, $n = \sum_{i=1}^{r} n_{i.}$,

$c = \sum_{i=1}^{r} c(i)$.

Für die gemeinsame Verteilung $F(x)$ der n Fehlervariablen gelte $F(x) = F(\pi x)$ für alle $\pi \in \times_{i=1}^{r} S_{n_{i.}}$. Man hat $\prod_{i=1}^{r} n_{i.}!$ Permutationen jeweils innerhalb der Stufen des Faktors A.

Getestet werden soll der Einfluß des genesteten Faktors B:

$$H: \beta_{1(i)} = \dots = \beta_{c(i)} \quad \text{für jedes } i=1,..,r$$

gegen

$$A: \beta_{j(i)} \neq \beta_{j'(i)} \quad \text{für ein } j(i)\neq j'(i)\in\{1,..,c(i)\} \text{ und ein } i\in\{1,..,r\} .$$

Dann hat man

$$U = \{\theta = \bigotimes_{i=1}^{r} \bigotimes_{j(i)}^{c(i)} \Delta_{ij}1_{n_{ij(i)}} \in R^n \mid \Delta_{ij(i)} = \mu+\alpha_i+\beta_{j(i)};\ \mu,\alpha_i,\beta_{j(i)}\in R, i=1,..,r; j(i)=1,..,c(i)\}$$

$$U^H = \{\theta\in U \mid \Delta_{ij(i)} = \mu+\alpha_i\} .$$

Unter H gilt:

$$\Delta_{i1} = \Delta_{i2} = \dots = \Delta_{ic(i)} \quad \text{für jedes } i=1,..,r$$

$$\Leftrightarrow \Delta_{ij(i)} - \bar{\Delta}_{i.} = 0$$

$$\Leftrightarrow K = I_n - \bigoplus_{i=1}^{r} 1/n_{i.} J_{n_{i.}} = \bigoplus_{i=1}^{r} (I_{n_{i.}} - 1/n_{i.} J_{n_{i.}}) .$$

Damit ist $S_n(K) = \times_{i=1}^{r} S_{n_{i.}}$.

Die transformierten Zufallsvariablen sind:

$$T_{ij(i)k} := (KY)_{ij(i)k} = Y_{ij(i)k} - \bar{Y}_{i..} .$$

Diese haben für jedes $i=1,..,r$ unter der Nullhypothese alle dieselbe Verteilung. Eine geeignete Teststatistik ist:

$$S(Y) = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} n_{ij} \bar{T}^2_{ij(i).} = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} n_{ij} (\bar{Y}_{ij(i).} - \bar{Y}_{i..})^2 .$$

$S(Y)$ ist die Summe von r Teststatistiken des unverbundenen k-Stichprobenproblems. Da $\bar{Y}_{i..}$ weiter invariant ist unter den zulässigen Permutationen jeweils innerhalb einer Stufe

des Faktors A, genügt es, folgende Teststatistik zu betrachten:

$$(3.48)\quad S(Y) = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} n_{ij} \bar{Y}^2_{ij(i)}.$$

Nur $\prod_{i=1}^{r} (n_{i.}! / \prod_{j(i)} n_{ij}!)$ Permutationen können zu unterschiedlichen Werten der Teststatistik führen.

Der Faktor A ist nur sinnvoll testbar, wenn man annimmt, daß die F_i - i=1,..,r - bis auf Lageverschiebungen identisch sind und der Faktor B keinen (signifikanten) Einfluß hat. Dann kann man einen Modellwechsel vornehmen und hat ein unverbundenes r-Stichprobenproblem mit $n_{i.}$ Beobachtungen pro Faktorstufe.

(viii) Weitere hierarchische und partiell-hierarchische Versuchspläne

Eine Übertragung des Tests aus (vii) auf einen $m \geq 2$-fach genesteten Versuchsplan (vgl. Kirk, 1968; S. 234f; Winer, 1971, S. 188f; Montgomery, 1976, Kap. 11.3) zur Feststellung eines Effekts des in der Hierarchie niedrigsten Faktors ist leicht möglich. Für das Modell des CRH-r(c)(s)

$$(3.49)\quad Y_{ij(i)k(i,j)l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(i,j)l} + \varepsilon_{ij(i)k(i,j)l}$$

mit i=1,..,r Stufen Faktor A,
j(i)=1,..,c(i) Stufen Faktor B genestet in A_i,
k(i,j)=1,..,s(i,j(i)) Stufen Faktor C genestet in $A_i, B_{j(i)}$,
l=1,..,n_{ijk} Replikationen

und $n_{ij.} = \sum_{k(i,j)} n_{ijk}$, $n = \sum_i \sum_{j(i)} \sum_{k(i,j)} n_{ijk}$

mit pro Stufen-Kombination $(A_i, B_{j(i)})$ symmetrisch verteilten Fehlern, d.h.

$$F_{ij(i)}(x) = F_{ij(i)}(\pi x) \quad \text{für alle } \pi \in S_{n_{ij.}}$$

findet man für den Test der Hypothese

(3.50) $H: \gamma_{1(i,j(i))} = \ldots = \gamma_{s(i,j(i))}$ für alle $i=1,..,r;$ $j(i)=1,..,c(i)$

gegen

$A: \gamma_{k(i,j(i))} \neq \gamma_{k'(i,j(i))}$ für ein $k \neq k' \in \{1,..,s(i,j(i))\}$ und ein $(i,j(i))$:

Die Transformationsmatrix ist:

$$K = \bigoplus_{i=1}^{r} \bigoplus_{j(i)=1}^{c(i)} (I_{n_{ij.}} - 1/n_{ij.} J_{n_{ij.}}) \quad \text{und} \quad S_n(K) = \times_{i=1}^{r} \times_{j(i)=1}^{c(i)} S_{n_{ij.}}$$

Die transformierten Zufallsvariablen sind

$$T_{ij(i)k(i,j)l} = Y_{ij(i)k(i,j)l} - \bar{Y}_{ij(i)..}$$

und die Teststatistik

$$S(T) = \sum_{i=1}^{r} \sum_{j(i)}^{c(i)} \sum_{k(i,j(i))}^{s(i,j(i))} n_{ijk} \bar{T}^2_{ij(i)k(i,j(i))k.}$$

oder äquivalent

(3.51) $$S(Y) = \sum_{i}^{r} \sum_{j(i)}^{c(i)} \sum_{k(i,j(i))}^{s(i,j(i))} n_{ijk} \bar{Y}^2_{ij(i)k(i,j(i))}.$$

Nur $\pi_i \pi_{j(i)} (n_{ij.}! / \pi_{k(i,j(i))} n_{ijk}!)$ Permutationen können zu unterschiedlichen Werten von S(Y) führen.

Analog zu (vii) kann nur bei Modellwechsel ein Test auf Effekte des in der Hierarchie nächsthöheren Faktors erfolgen.

Genauso direkt erfolgt die Erweiterung auf den partiell-hierarchischen CRH-(rc)(s) - Plan, in dem ein Faktor C unter die Faktorstufenkombinationen der beiden gekreuzten Faktoren A und B genestet ist. Das Modell lautet:

(3.52) $$Y_{ijk(i,j)l} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(i,j)} + \varepsilon_{ijk(i,j)l}$$

(Indizes in Analogie zu (3.49)!)

mit je Faktorstufenkombinationen $A_i \times B_j$ symmetrisch verteilten

$$\varepsilon_{ij} = (\varepsilon_{ij11}, \ldots, \varepsilon_{ijs(i,j)n_{ijs(i,j)}})'$$

mit Verteilungsfunktionen

$$F_{ij}(x) = F_{ij}(\pi x) \quad \text{für alle } \pi \in S_{n_{ij}} .$$

Zum Testen der Hypothese (3.50) mit $c(1)=\ldots=c(r)=c$ läßt sich in Analogie zu (3.51) direkt eine Teststatistik angeben:

$$(3.53) \quad S(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k(i,j)}^{s(i,j)} n_{ijk} \bar{Y}^2_{ijk(i,j)} .$$

Falls keine Effekte des genesteten Faktors vorhanden sind, kann nach Modellwechsel (d.h. (3.52) ohne den Effekt $\gamma_{k(i,j)}$) wegen der Ergebnisse für den CRF-rc-Plan kein Invarianz-Permutationstest auf Wechselwirkung zwischen den Faktoren A und B angegeben werden.

Für das zweite mögliche partiell-hierarchische Modell mit drei Faktoren CRH-r(c)s (vgl. Kirk, 1968, S. 235f; Winer, 1971, S. 186ff) ist eine Übertragung dieser Überlegungen nicht möglich. In diesem Modell ist Faktor B unter Faktor A genestet, jedoch gekreuzt mit Faktor C:

$$(3.54) \quad Y_{ij(i)kl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{j(i)k} + \varepsilon_{ij(i)kl}$$

$i=1,..,r$ Stufen Faktor A,
$j(i)=1,..,c(i)$ Stufen Faktor B unter A_i
$k=1,..,s$ Stufen Faktor C,
$l=1,..,n_{ijk}$ Replikationen

mit je Faktorstufe A_i symmetrisch verteilten

$$\varepsilon_i = (\varepsilon_{i111}, \ldots, \varepsilon_{ic(i)sn_{ic(i)s}})' \text{ nach } F_i(x) = F_i(\pi x) \text{ für alle } \pi \in S_{n_{i..}}$$

Pro Faktorstufe A_i handelt es sich um einen zweifaktoriellen Zufallsplan CRF-c(i)s. Nimmt man o.B.d.A. $c(i)=c$ und $n_{ijk} = n$ an, hat man mit (3.30) als Transformationsmatrix K zum Testen der Hypothese

$$(3.55) \quad H: \quad (\beta\gamma)_{1(1)1} = \ldots = (\beta\gamma)_{c(r)s} = 0$$

$$K = I_r \otimes (I_{csn} - 1/sn I_c \otimes J_{sn} - 1/cn J_c \otimes I_s \otimes J_n + 1/csn J_{csn})$$

und

$$(KY)_{ij(i)kl} := T_{ij(i)kl} = Y_{ij(i)kl} - \bar{Y}_{ij(i)..} - \bar{Y}_{i.k.} + \bar{Y}_{i...} \quad .$$

Jede sinnvolle Teststatistik basiert wieder auf den Zellmitteln oder -summen:

$$(3.56) \quad S(Y) = \sum_i \sum_{j(i)} \sum_k T_{ij(i)k.} = \sum_i \sum_{j(i)} \sum_k n(\bar{Y}_{ij(i)k.} - \bar{Y}_{ij(i)..} - \bar{Y}_{i.k.} + \bar{Y}_{i...})$$

S ist aber invariant gegen Vertauschungen von Elementen innerhalb der Zellen, von ganzen Zeilen und von ganzen Spalten je Faktorstufe A_i und von ganzen 'Teilversuchsplänen' je Stufe des Faktors A. Damit ist ein Invarianz-Permutationstest für die Hypothese (3.55) nicht möglich.
In Analogie zum zweifaktoriellen Zufallsplan ist ein Invarianz-Permutationstest auf die Wechselwirkung der beiden nicht genesteten Faktoren A und C ebenfalls unmöglich.

(ix) Lateinische Quadrate (LS-k; Welch, 1937; Scheffe, 1959; Pyhel, 1980b; Willmes, 1981b)

In einem Lateinischen Quadrat können die Einflüsse von zwei Störfaktoren (A, B) bei der Beurteilung eines interessierenden (Behandlungs-) Effektes (C) minimiert werden (vgl. Scheffe, 1959, Kap. 9.2, Fall (iii); Kirk, 1968, Kap. 6.1). Anwendungen in der Psychologie findet man in Bortz (1977, Kap. 11.2) oder Winer (1971, S. 522f). Die Faktoren A, B, C, besitzen jeweils l Faktorstufen. Es werden insgesamt nur l^2 Beobachtungen durchgeführt, so daß in jeder 'Zeile' und 'Spalte' des Versuchsplans jede Faktorstufe von C genau einmal auftritt. Unter Annahme der Additivität der drei Faktoren lautet das Modell:

$$(3.57) \quad Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{k(i,j)} + \varepsilon_{ij} \qquad \begin{matrix} i,j=1,..,l, \\ k(i,j') \neq k(i,j) \neq k(i',j) \end{matrix}$$

mit den Reparametrisierungsbedingungen $\alpha_\cdot = \beta_\cdot = \gamma_\cdot = 0$
und $\varepsilon = (\varepsilon_{11},..,\varepsilon_{ll})'$ verteilt nach symmetrischem $F(x) = F(\pi x)$ für alle $\pi \in S_{l^2}$.

Es interessiert die Hypothese:

$$H: \quad \gamma_1 = \ldots = \gamma_l$$

gegen

$$A: \quad \gamma_k \neq \gamma_{k'} \quad \text{für ein } k \neq k' \in \{1,..,l\} \quad .$$

In der üblichen Schreibweise mit

$$U = \{\theta \in R^{l^2} \mid \theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{k(i,j)};\ \mu, \alpha_i, \beta_j, \gamma_{k(i,j)} \in R; i,j,k = 1,..,l\}$$

hat man

$$U^H = \{\theta \in U \mid \theta_{ij} = \mu + \alpha_i + \beta_j\} \quad .$$

Zur Bestimmung von K überlegt man:

$$\begin{aligned} U^H &\Longleftrightarrow \theta_{ij} = \bar{\theta}_{i.} + \bar{\theta}_{.j} - \bar{\theta}_{..} \\ &\Longleftrightarrow \theta_{ij} - \bar{\theta}_{i.} - \bar{\theta}_{.j} + \bar{\theta}_{..} = 0 \\ &\Longleftrightarrow K\theta = 0 \end{aligned}$$

mit idempotentem $K = (I_l - 1/l J_l) \otimes (I_l - 1/l J_l)$.

Damit hat man

$$(KY)_{ij} = T_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

mit folgenden zulässigen Permutationen (vgl. Pyhel, 1978a, S. 11):

$$Q_{l^2} = S_{l^2}(K) \cap S_{l^2} = S_l \otimes S_l \cap S_{l^2} = S_l \otimes S_l \quad .$$

Q_{l^2} ist die Menge aller $(l!)^2$ Zeilen- und Spaltenpermutationen in dem Versuchsplan.

Als Teststatistik benutzt man

$$S(T) = \sum_{k=1}^{l} T_k^2 \quad \text{mit} \quad T_k = \sum_{k(i,j)=k} T_{ij} \quad (\text{oder } T_k = 1/l \sum_{k(i,j)=k} T_{ij}) \quad . \tag{3.58}$$

Zur Durchführung des Permutationstests werden alle l^2 Zeilen- und Spaltenpermutationen der transformierten t_{ij}, $i,j=1,..,l$, ausgeführt für ein zufällig ausgewähltes Lateinisches Quadrat. Bei dieser Zufallsauswahl ist folgendes zu beachten. Es wird ein Quadrat zufällig aus einer sog. Transformationsmenge ausgewählt. Diese enthält alle Lateinischen Quadrate, die durch Vertauschungen von Zeilen, Spalten oder Faktorstufen von Faktor C in den Zellen entstehen. Eine

Transformationsmenge läßt sich charakterisieren durch die in ihr enthaltenen Standardquadrate. Diese haben die Eigenschaft, in der ersten Zeile und Spalte des Lateinischen Quadrates die Faktorstufen von C der Große nach aufsteigend geordnet zu enthalten. Die Gesamtzahl an Quadraten in einer Transformationsmenge ist $l!\cdot(l-1)!\cdot s_l$, wobei s_l die Anzahl Standardquadrate bezeichnet. Da die Teststatistik (3.58) invariant ist gegen die l! Vertauschungen der Stufen von C, gibt es $(l-1)!s_l$ mogliche verschiedene Werte von S. In Fisher & Yates (1963) ist die Anzahl Transformationsmengen mit der zugehörigen Anzahl von Standardquadraten für kleine l angegeben:

l	Anz.Transf.Mengen t	Standardquadrate	Max.Anz.Werte der Teststatistik	$l!^2$
3	1	1	2	36
4	2	3,1	18,6	576
5	2	50,6	1200,144	14400
6	12	max=1080,min=20	max=129600,min=2400	518400

Für l=4,5 kann sich damit die übliche Empfehlung von Fisher & Yates (vgl. Kirk, 1968, S. 154), zufällig eines der Standardquadrate auszuwählen, deutlich auf die möglichen α-Niveaus für den exakten Permutationstest auswirken. Zwei Vorgehensweisen sind möglich:

(i) Man wählt grundsätzlich ein Quadrat aus der größeren Transformationsmenge aus.

(ii) Man wählt per Zufall ein Standardquadrat. Nach Bestimmung aller möglichen verschiedenen Werte von S ordnet man die t_{ij} nach einem Standardquadrat aus der jeweils

anderen Transformationsmenge und bestimmt die übrigen möglichen Werte von S. Bei zufälliger Auswahl des Standardquadrates ist dieses Vorgehen erlaubt, denn bei Gültigkeit der Nullhypothese sind alle $l! \cdot \prod_{i=1}^{t} s_{1_i}$ Werte von S gleichwahrscheinlich (vgl. Nelder, 1964b, S. 168; aber auch Pearson, 1937, S. 62ff).

Für $l \geq 6$ kann ein Standardquadrat zufällig ausgewählt werden (Wu, 1981). Eine Erweiterung des Versuchsplans auf $m>1$ Replikationen je Zelle des Versuchsplans ist leicht möglich. Das Modell lautet in diesem Fall:

$$Y_{ijh} = \mu + \alpha_i + \beta_j + \gamma_{k(i,j)} + \varepsilon_{ijh} \quad \text{mit } h=1,..,m>1$$

und

$$\varepsilon = (\varepsilon_{111}, \ldots, \varepsilon_{11m})' \text{ verteilt nach symmetrischem } F(x)=F(\pi x)$$
$$\text{für alle } \pi \in S_{l^2 m} \quad .$$

Zum Test der obigen Hypothese benötigt man

$$K^* = K \otimes 1/m J_m \quad \text{und als zulässige Permutationsgruppe}$$
$$Q_{l^2 m} = S_{l^2 m}(K^*) = S_m \wr S_{l^2}(K) = S_m \wr (S_l \otimes S_l) \quad ,$$

d.h. es kommen nur noch die Teststatistik

$$S^*(T) = \sum_{k=1}^{l} \bar{T}_{k.}^2 \quad \text{mit} \quad \bar{T}_{k.} = 1/m \sum_{k(i,j)=k}^{\cdot} \sum_{h=1}^{m} T_{ijh}$$

invariant lassende $m!^{l^2}$ Vertauschungen innerhalb der Zellen des Lateinischen Quadrates hinzu.

(x) 2-faktorieller vollständiger Zufallsplan (CRF - rc)

Wie zu Beginn dieses Abschnittes gezeigt, existiert kein Invarianz-Permutationstest auf Wechselwirkungen. Ein Test der Hypothese: 'Der Faktor B (bzw. A) hat überhaupt keinen Einfluß, d.h. auf keiner Stufe des Faktors A (bzw. B)' kann dagegen angegeben werden (vgl. Kendall & Stuart, 1968, S.106 für den Fall einfacher Zellbesetzung!). Im Gegensatz zu 3.22 lautet die Modellgleichung:

(3.59) $Y_{ijk} = \mu + \beta_{ij} + \varepsilon_{ijk}$ $\quad i=1,..,r;\ j=1,..,c;\ k=1,..,n_{ij}>1;$

$$n=\sum_{ij}\sum n_{ij}$$

mit den Reparametrisierungsbedingungen

$$\beta_{i.}=0 \quad \text{für alle } i=1,..,r$$

und

$$\varepsilon=(\varepsilon_{111},...,\varepsilon_{rcn_{rc}})' \text{ verteilt nach } F(x)=F(\pi x)$$

für alle $\pi \in S_n$.

Die zu testende Hypothese lautet:

$$H: \quad \beta_{i1} = ... = \beta_{ic} \quad \text{für jedes } i=1,..,r$$

gegen

$$A: \quad \beta_{ij} \neq \beta_{ij'} \quad \text{für ein } j \neq j' \in \{1,..,c\} \text{ und } i \in \{1,..,r\} \ .$$

Damit hat man - bis auf die etwas andere Forderung an die Fehler - dieselbe Situation wie im einfach hierarchischen Zufallsplan.

Die Transformationsmatrix ist deshalb:

$$K = \bigotimes_{i=1}^{r} (I_{n_{i.}} - 1/n_{i.} J_{n_{i.}})$$

mit den transformierten Zufallsvariablen $T_{ijk}=Y_{ijk}-\bar{Y}_{i..}$ und den zulässigen Permutationen $Q_n = \times_{i=1}^{r} S_{n_{i.}}$, d.h. unabhängigen Vertauschungen innerhalb jeder Stufe des Faktors A. Eine geeignete Teststatistik ist dann

$$S(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \bar{T}^2_{ij.}$$

oder, da $\bar{Y}_{i..}$ invariant ist unter Ω_n,

(3.60) $$S(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \bar{Y}^2_{ij.} \quad .$$

Mit derselben Teststatistik testet man analog:

$$H: \alpha_{1j} = \ldots = \alpha_{rj} \quad \text{für alle } j=1,..,c \quad .$$

In diesem Fall werden jedoch alle zulässigen Permutationen aus $\Omega_n = \times_{j=1}^{c} S_{n_{.j}}$, d.h. unabhängige Vertauschungen innerhalb jeder Stufe von Faktor B, betrachtet.

(xi) Split-Plot Plan, Profilparallelität (SPF-r.c; Pyhel, 1978a, b, 1980a; Willmes & Pyhel, 1981; Willmes, 1982; Stein-Zaar, 1983)

Dieser Versuchsplan wird u.a. angewendet, wenn (a) $n_{.} = \sum_{i=1}^{r} n_i$ Beobachtungseinheiten (Vpn) r Bedingungen eines Faktors A (Plot-Faktor), für die i.a. keine wiederholten Messungen vorgenommen werden können, zufällig zugewiesen werden und wenn jede Beobachtungseinheit in zufälliger Reihenfolge unter c Stufen eines (Split-) Faktors B analysiert wird (Kirk, Kap. 8.1; Koch et al., 1980). Es soll dann getestet werden, ob Wechselwirkungen zwischen beiden Faktoren vorliegen.

Eine andere typische Situation ergibt sich, wenn (b) der Faktor B verschiedene, fixierte Zeitpunkte darstellt, sodaß pro Einheit ein 'Profil' von Beobachtungen vorliegt. Im letzteren Fall ist von Interesse, ob die experimentellen Bedingungen des Faktors A auf den einzelnen Zeitstufen zu verschiedenen Effekten führen, oder ob die Zeiteffekte unter allen experimentellen Bedingungen gleich sind, d.h. ob die

Profile für alle Stufen von A 'parallel' verlaufen (vgl. Collier, Baker, & Mandeville, 1967).

Das Modell im Fall (a) lautet:

$$Y_{ih(i)j} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\pi_{h(i)}+\varepsilon_{ih(i)j} \quad \begin{matrix} i=1,..,r \\ h(i)=1,..,n_i \\ j=1,..,c;\ n_\cdot=\sum_{i=1}^{r} n_i \\ N=n_\cdot c \end{matrix}$$

(3.61) mit $\pi_{h(i)}$ fester Effekt des h-ten Blockes der h-ten Beobachtungseinheit unter der Faktorstufe A_i und den Reparametrisierungsbedingungen

$$\alpha_\cdot=\beta_\cdot=(\alpha\beta)_{i\cdot}=(\alpha\beta)_{\cdot j}=(\alpha\beta)_{\cdot\cdot}=\pi_{\cdot(i)}=0 \quad .$$

$\varepsilon=(\varepsilon_{111},\ldots,\varepsilon_{rn_rc})'$ sei verteilt nach $F(x)$, $x\in R^N$;

mit $F(\pi x)=F(x)$ für alle $\pi\in S_c\sim S_{n_\cdot}$.

Die 'Fehler-Blöcke' $\varepsilon_{ih(i)}=(\varepsilon_{ih(i)1},\ldots,\varepsilon_{ih(i)c})'$ haben vertauschbare Komponenten, und sie können über alle $n_\cdot$ Blöcke hinweg vertauscht werden. Diese Bedingung impliziert einerseits, daß alle Varianzen und Kovarianzen der $\varepsilon_{ih(i)j}$ je Block gleich sind. Andererseits dürfen die Blöcke keine Variabilitätsunterschiede unter verschiedenen Stufen von A aufweisen, da die Vertauschbarkeit über verschiedene Stufen von A hinweg dann verletzt wäre. Insgesamt enthält $S_c\sim S_{n_\cdot}$ $n_\cdot!c!^{n_\cdot}$ Permutationen (vgl. (3.20a)). Getestet werden soll die Hypothese:

$$H: \ (\alpha\beta)_{11}=\ldots=(\alpha\beta)_{rc}=0$$

(3.62) gegen

$$A: \ (\alpha\beta)_{ij}\neq(\alpha\beta)_{i'j'} \text{ für ein } (i,j)\neq(i',j') \quad .$$

Im Fall (b) hat man als Modell

$$Y_{ih(i)j}=\mu_{ih(i)j}+\varepsilon_{ih(i)j} \quad \text{mit i,h(i),j wie zuvor}$$

mit den $n_\cdot$ 'Profilen' $\mu_{ih(i)}=(\mu_{ih(i)1},\ldots,\mu_{ih(i)c})'$.

Hier wird für die Fehler-Blöcke nur angenommen, daß sie über alle $n_.$ Blöcke vertauschbar sind, d.h. $F(x) = F(\pi x)$ für alle $\pi \in \hat{S}_{n_.} = S_{n_.} \otimes \{\iota_c\}$. $\hat{S}_{n_.}$ enthält $n_.!$ Permutationen von c-dimensionalen Vektoren. Es sind im folgenden auch Niveauunterschiede - jedoch wie oben keine Variabilitätsunterschiede - zwischen den Blöcken zugelassen. Durch $\hat{S}_{n_.}$ ist nicht ausgeschlossen, daß es für die verschiedenen Split-Stufen unterschiedliche Variabilitäten gibt.

Die Hypothese der Profilparallelität für alle $n_.$ Profile gegen die Alternative, daß innerhalb jeder Plotstufe die n_i Profile zwar parallel, aber für mindestens zwei Plotstufen i und i' die Profile nicht parallel seien, läßt sich folgendermaßen formalisieren:

$$H: \text{ es gibt } (\Delta_1,..,\Delta_{c-1})' \in R^{c-1} \text{ mit}$$

$$(3.63) \qquad \mu_{ih(i)j} - \mu_{ih(i)j+1} = \Delta_j \quad \text{für alle } \begin{matrix} j=1,..,c-1; \\ i=1,..,r; \\ h(i)=1,..,n_i \end{matrix}$$

gegen

$$A: \mu_{ih(i)j} - \mu_{ih(i)j+1} = \Delta_{ij} \text{ für alle } h(i)=1,..,n_i$$

oder äquivalent (vgl. Pyhel, 1980a):

$$H: \mu_{ih(i)j} - \bar{\mu}_{ih(i).} - \bar{\mu}_{..j} + \bar{\mu}_{...} = 0$$

gegen

$$A: \mu_{ih(i)j} - \bar{\mu}_{ih(i).} - \bar{\mu}_{..j} + \bar{\mu}_{...} = \nu_{ij} \neq 0$$

für ein Paar (i,j) und alle $h(i)=1,..,n_i$.

In Matrixschreibweise bedeutet das - mit $\mu = (\mu'_{11},..,\mu'_{rn_r})'$ -

$$H: K\mu = 0 \quad \text{gegen} \quad A: K\mu \neq 0 \quad \text{mit}$$

$$(3.64) \qquad K = (I_N - 1/cI_{n_.} \otimes J_c - 1/n_. J_{n_.} \otimes I_c + 1/NJ_N)$$

$$= (I_{n_.} - 1/n_. J_{n_.}) \otimes (I_c - 1/cJ_c) \quad .$$

Im Fall (a) hat man:

$$U = \{\theta \in R^N \mid \theta_{ih(i)j} = \mu + \alpha_i + \beta_j + \pi_{h(i)} + (\alpha\beta)_{ij};\ \mu, \alpha_i, \beta_j, \pi_{h(i)}, (\alpha\beta)_{ij} \in R;\ i=1,..,r; h(i)=1,..,n_i; j=1,..,c\}$$

und

$$U^H = \{\theta \in U \mid \theta_{ih(i)j} = \mu + \alpha_i + \beta_j + \pi_{h(i)}\} \quad .$$

Es gilt $\theta \in U^H \Leftrightarrow \theta_{ih(i)j} - \bar{\theta}_{ih(i).} - \bar{\theta}_{..j} + \bar{\theta}_{...} = 0$.

Damit hat man dieselbe Transformationsmatrix K wie in Fall (b).

Angewendet auf Y ergibt sich:

$$(KY)_{ih(i)j} := T_{ih(i)j} = Y_{ih(i)j} - \bar{Y}_{ih(i).} - \bar{Y}_{..j} + \bar{Y}_{...} \quad .$$

Mit den Regeln von Pyhel (1978a, S. 11) hat man sofort:

$$S_N(K) = S_{n.} \otimes S_c \quad ,$$

d.h. Vertauschungen aller n. c-dimensionalen Blöcke untereinander und der c Stufen des Split-Faktors B über alle Blöcke hinweg. Im Fall (a) ergibt sich $\mathcal{Q}_N$ nach (3.14) als $\mathcal{Q}_N = S_N(K)$, denn $S_N(K)$ ist Teilmenge von $S_c \wr S_{n.}$; dagegen hat man im Fall (b) $\mathcal{Q}_N = S_{n.} \otimes \{\iota_c\}$.

Zur Ermittelung einer geeigneten Teststatistik muß K_A bestimmt werden. Mit den Reparametrisierungsbedingungen gilt:

$$\bar{\theta}_{i.j} - \bar{\theta}_{i..} - \bar{\theta}_{..j} + \bar{\theta}_{...} = (\alpha\beta)_{ij}$$

oder in Matrixschreibweise:

$$K_A = \bigoplus_{i=1}^{r} (1/n_i J_{n_i} \otimes I_c) - \bigoplus_{i=1}^{r} (1/n_i J_{n_i} \otimes 1/cJ_c) - 1/n_{.} J_{n_{.}} \otimes (I_c - 1/cJ_c$$

$$= \bigoplus_{i=1}^{r} (1/n_i J_{n_i} \otimes (I_c - 1/cJ_c)) - 1/n_{.} J_{n_{.}} \otimes (I_c - 1/cJ_c) \quad .$$

Durch Ausrechnen findet man:

$$K_A(T) = (\bar{T}_{1.1}, \ldots, \bar{T}_{1.1}, \bar{T}_{1.2}, \ldots, \bar{T}_{1.2}, \ldots, \bar{T}_{r.c}, \ldots, \bar{T}_{r.c})' \quad ,$$

(n_1 - mal, n_2 - mal, n_r - mal)

da

$$\bar{T}_{ih(i).} = \bar{T}_{i..} = \bar{T}_{..j} = \bar{T}_{...} = 0 \quad \text{für alle } i, h(i), j \quad .$$

Folglich hat man als Teststatistik:

$$S(Y) = \sum_{i=1}^{r} n_i \sum_{j=1}^{c} \bar{T}^2_{i.j} \tag{3.65}$$

$$= \sum_{i=1}^{r} n_i \sum_{j=1}^{c} (\bar{Y}_{i.j} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})^2 \quad .$$

Da die $\bar{Y}_{..j}$ und $\bar{Y}_{...}$ invariant - bis auf Umindizierungen der Splitfaktorstufen in Fall (a) unter den zulässigen Permutationen sind, genügt es, folgende Teststatistik zu betrachten:

$$S(Y) = \sum_{i=1}^{r} n_i \sum_{j=1}^{c} (\bar{Y}_{i.j} - \bar{Y}_{i..})^2 \quad . \tag{3.66}$$

In dieser Form zeigt sich der Zusammenhang zum RB-k: In (3.65) hat man die Summe von r mit den jeweiligen Stichprobenumfängen gewichteten RB-c Statistiken. Der zusätzliche Index i kennzeichnet die verschiedenen Plot-Stufen. Die zulässigen Permutationen entsprechen - angewendet jeweils auf die ganzen c-dimensionalen Blöcke - denen im CR-r Plan. Die Vertauschungen von Split-Stufen über alle Blöcke hinweg gemäß S_c verändern S(Y) nicht. Nur die $n_{\cdot}!/\prod_{i=1}^{r} n_i!$ Vertauschungen von Blöcken zwischen verschiedenen Plot-Stufen können zu unterschiedlichen Werten der Teststatistik führen. Koch hat ähnliche (Permutations-) Rangtests für die Fälle (a) und (b) angegeben (1969, S. 497 unten bzw. 495, Formel (34) und (35)).

Bisher war stets die euklidische Länge von $f_A(T)$ als Teststatistik verwendet worden. Prinzipiell ist jedoch auch eine quadratische Form in $f_A(T)$ als Teststatistik geeignet. Diese soll im folgenden für beide Fälle (a) und (b) vorgestellt werden.

Fall (a): Bei gegebener Realisation t von T=KY haben, wenn man die $T_{ih(i)j}$ genauso wie die $Y_{ih(i)j}$ in das Versuchsschema einordnet, alle Zeilen- und Spaltenvertauschungen für t dieselbe Wahrscheinlichkeit $P_N = 1/(n_\cdot!c!)$. Das ist der reziproke Wert der Anzahl der Elemente von $S_N(K)$. Mathematisch exakt ist P_N bei Pyhel (1978a, S.59) definiert.
Damit gilt für die ein- und zweidimensionalen Randverteilungen von T bei gegebenem t:

$$P_N(T_{ih(i)j}=x) = (n_\cdot-1)!(c-1)!/n_\cdot!c! = 1/N = 1/n_\cdot c$$

$$P_N(T_{ih(i)j}=x,\ T_{i'h'(i')j'}=x') = \begin{cases} 1/(n_\cdot c(c-1)) & (i,h(i))=(i',h'(i')) \\ & j\neq j' \\ 1/(n_\cdot(n_\cdot-1)c)) & (i,h(i))\neq(i',h'(i')) \\ & j=j' \\ 1/(n_\cdot(n_\cdot-1)c(c-1)) & (i,h(i))\neq(i',h'(i')) \\ & j\neq j' \end{cases}$$

Damit können Erwartungswerte, Varianzen und Kovarianzen der $T_{ih(i)j}$ berechnet werden:

$$E_{P_N}(T_{ih(i)j}) = 1/n_\cdot c \sum_{i,h(i),j} t_{ih(i)j} = \bar{t}_{\cdots} = 0$$

$$Var_{P_N}(T_{ih(i)j}) = 1/n_\cdot c \sum_{i,h(i),j} t^2_{ih(i)j} =: (n_\cdot-1)(c-1)/n_\cdot c\sigma^2_{N,t}$$

$$\text{mit } \sigma^2_{N,t} = 1/(n_\cdot-1)(c-1) \sum_{i,h(i),j} t^2_{ih(i)j} \quad .$$

Für die drei oben unterschiedenen Fälle findet man nach längeren Umformungen für die Kovarianzen:

$$Cov_{P_N}(T_{ih(i)j}, T_{ih(i)j'}) = -(n_\cdot-1)/n_\cdot c\sigma^2_{N,t}$$

$$Cov_{P_N}(T_{ih(i)j}, T_{i'h'(i')j}) = -(c-1)/n_\cdot c\sigma^2_{N,t}$$

$$Cov_{P_N}(T_{ih(i)j}, T_{i'h'(i')j'} = 1/n_\cdot c\sigma^2_{N,t}$$ (Schreibfehler in Pyhel (1978a,S.61 !)).

Zur Konstruktion der quadratischen Form müssen die entsprechenden Größen für die $\bar{T}_{i\cdot j}$ aus obigen Ergebnissen ermittelt werden:

$$E_{P_N}(\bar{T}_{i.j}) = 0$$

$$Var_{P_N}(\bar{T}_{i.j}) = (n_{.}-n_i)(c-1)/n_i n_{.} c\sigma^2_{N,t} = 1/n_{.}(n_{.}/n_i - 1)(1-1/c)\sigma^2_{N,t}$$

$$Cov_{P_N}(\bar{T}_{i.j},\bar{T}_{i.j'}) = -(n_{.}-n_i)/n_i n_{.} c\sigma^2_{N,t} = 1/n_{.}(n_{.}/n_i - 1)(-1/c)\sigma^2_{N,t}$$

$$Cov_{P_N}(\bar{T}_{i.j},\bar{T}_{i'.j}) = -(c-1)/n_{.} c\sigma^2_{N,t} = 1/n_{.}(\quad -1)(1-1/c)\sigma^2_{N,t}$$

$$Cov_{P_N}(\bar{T}_{i.j},\bar{T}_{i'.j'}) = 1/n_{.} c\sigma^2_{N,t} = 1/n_{.}(\quad -1)(-1/c)\sigma^2_{N,t} .$$

Mit $T_N = (\bar{T}_{1.1},\ldots,\bar{T}_{1.c},\ldots,\bar{T}_{r.1},\ldots,\bar{T}_{r.c})'$ hat man dann:

$$E_{P_N}(T_N) = 0$$

$$V_N := Var_{P_N}(T_N) = E_{P_N}(T_N \cdot T_N') = ((n_{.}/n_i\delta_{ii'} - 1)) \otimes (I_c - 1/cJ_c)\cdot\sigma^2_{N,t}/n_{.}$$

V_N ist nicht von vollem Rang sondern hat Rang $(r-1)(c-1)$. Deshalb muß zur Bildung der quadratischen Form eine verallgemeinerte Inverse (vgl. Anhang A2) gebildet werden. Diese ist, da $(I_c - 1/cJ_c)$ idempotent und damit identisch mit seiner verallgemeinerten Inversen ist:

$$V_N^- = (n_{.}/\sigma^2_{N,t})\cdot((n_{.}/n_i\delta_{ii'} - 1))^- \otimes (I_c - 1/cJ_c)$$
$$= (n_{.}/\sigma^2_{N,t})\cdot((n_i/n_{.}\delta_{ii'} \pm n_i n_{i'}/n_r)) \otimes (I_c - 1/cJ_c)$$

(vgl. Puri & Sen, 1971, S. 185, Formel (5.4.11)).

Die quadratische Form ist dann:

$$(3.67)\quad Q(T_N) = T_N' V_N^- T_N = 1/\sigma^2_{N,t} \sum_{i=1}^{r}\sum_{j=1}^{c} n_i \bar{T}^2_{i.j}$$
$$= S(T)/\sigma^2_{N,t} \quad \text{(vgl. (3.65))} .$$

Da $\sigma^2_{N,t}$ konstant ist über alle zulässigen Vertauschungen, sind Q und S äquivalent (vgl. (3.65)).

Im Fall gleicher Stichprobenumfänge $n_i=n$ (Pyhel, 1978a) hat man:

$$V_N = (\sigma^2_{N,t}/n)\cdot(I_r - 1/rJ_r)\otimes(I_c - 1/cJ_c) \quad .$$

Dann ist $V_N^- = n/\sigma^2_{N,t} I_{rc}$ eine verallgemeinerte Inverse und

$$Q(T_N) = 1/\sigma^2_{N,t} \sum_{i=1}^{r} \sum_{j=1}^{c} n\bar{T}^2_{i.j} \quad .$$

Fall (b): Jetzt haben alle Vertauschungen von Zeilen im Versuchsschema für die $t_{ih(i)j}$ dieselbe Wahrscheinlichkeit $1/n_.!$:

$$P_{n.}(T_{ih(i)j}=x) = (n_.-1)!/n_.! = 1/n_. \quad \text{(Exakte Definition analog zu Pyhel (1978a,S.59)}$$

$$P_{n.}(T_{ih(i)j}=x,\ T_{i'h'(i')j'}=x') = \begin{cases} 1/n_. & (i,h(i))=(i',h'(i')) \\ & j\neq j' \\ 1/(n_.(n_.-1)) & (i,h(i))\neq(i',h'(i')) \\ & j,j' \text{ beliebi} \end{cases}$$

Erwartungswerte, Varianzen und Kovarianzen bestimmt man zu (vgl. Phyel, 1980a):

$$E_{P_{n.}}(T_{ih(i)j}) = 1/n_. \sum_{i,h(i)} t_{ih(i)j} = \bar{t}_{..j} = 0$$

$$Var_{P_{n.}}(T_{ih(i)j}) = 1/n_. \sum_{i,h(i)} t^2_{ih(i)j} =: (n_.-1)/n_.\cdot\sigma_{jj}$$

mit $\sigma_{jj} := 1/(n_.-1) \sum_{i,h(i)} t^2_{ih(i)j}$ je Profilstufe verschieden

$$Cov_{P_{n.}}(T_{ih(i)j}, T_{ih(i)j'}) = 1/n_. \sum_{i,h(i)} t_{ih(i)j}\cdot t_{ih(i)j'} =: (n_.-1)/n_.\sigma$$

mit $\sigma_{jj'} := 1/(n_.-1) \sum_{i,h(i)} t_{ih(i)j}\cdot t_{ih(i)j'}$ verschieden je Pa
von Profilstufen im Unterschied zu Fall (a) .

$$Cov_{P_{n.}}(T_{ih(i)j}, T_{i'h'(i')j}) = -1/n_.\sigma_{jj}$$

$$Cov_{P_{n.}}(T_{ih(i)j}, T_{i'h'(i')j'}) = -1/n_.\sigma_{jj'} \quad .$$

Zur Konstruktion der quadratischen Form benötigt man wieder (vgl. Pyhel, 1980a):

$$E_{P_{n.}}(\bar{T}_{i.j}) = 0 \quad ; \quad Var_{P_{n.}}(\bar{T}_{i.j}) = 1/n_{.}(n_{.}/n_{i}-1)\sigma_{jj}$$

$$Cov_{P_{n.}}(\bar{T}_{i.j},\bar{T}_{i.j'}) = 1/n_{.}(n_{.}/n_{i}-1)\sigma_{jj'}$$

$$Cov_{P_{n.}}(\bar{T}_{i.j},\bar{T}_{i'.j}) = -1/n_{.}\sigma_{jj} \quad ; \quad Cov_{P_{n.}}(\bar{T}_{i.j},\bar{T}_{i'.j'}) = -1/n_{.}\sigma_{jj'} \; .$$

Damit gilt: $E_{P_{n.}}(T_N) = 0$; $V_{n.} := E_{P_{n.}}(T_N \cdot T_N')$

und mit $((\sigma_{jj'}))=: \Sigma_{n.}$: $= 1/n_{.}((n_{.}/n_{i}\delta_{ii'}-1)) \otimes ((\sigma_{jj'}))_{c\times c}$

$$V_{n.}^{-} = n_{.}((n_{i}/n_{.}\delta_{ii'} \pm n_{i}n_{i'}/n_{r})) \otimes \Sigma_{n.}^{-} \quad \text{mit} \quad \Sigma_{n.}^{-} =: ((\sigma^{*}_{jj'}))_{c\times c} \; .$$

Als Teststatistik erhält man dann die quadratische Form (vgl. Pyhel, 1980a):

$$(3.68) \quad Q(T_N) = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{j'=1}^{c} n_i \bar{T}_{i.j} \sigma^{*}_{jj'} \bar{T}_{i.j'} \; .$$

$((\sigma^{*}_{jj'}))$ ist invariant unter den zulässigen Vertauschungen der jeweils ganzen Blöcke. Die (möglicherweise verallgemeinerte) Inverse muß nur einmal zur Bestimmung des $p^{\geq}$-Wertes berechnet werden. $Q(T_N)$ nimmt höchstens $n_{.}! / \prod_{i=1}^{r} n_i!$ verschiedene Werte an. Der Rechenaufwand ist verglichen mit (3.65) also nicht wesentlich größer. Allerdings ist das Ergebnis, falls $\Sigma_{n.}^{-}$ nicht von vollem Rang sein sollte, abhängig von der speziellen Wahl der verallgemeinerten Inversen. Eine eindeutige Festlegung könnte man durch Wahl der Moore-Penrose Inverse (vgl. Pyhel, 1980a, Kap. 8) erzielen.

Pyhel (1980a) gibt an, daß Q nur für 'große' Stichproben (d.h. für den asymptotischen Fall) verwendet werden solle. In diesem Fall wird dann nicht die exakte Verteilung von $Q(T_N)$ bestimmt, sondern ausgenutzt, daß Q für $n_{.} \to \infty$ gegen eine χ^2_f- Verteilung mit $(r-1) \leq f \leq (r-1)(c-1)$ konvergiert

(Details siehe Pyhel, 1980a, Abschnitt 7). Nimmt man wie im Fall (a) an, daß die ε_{ih} S_c-invariant sind, was Identität der Varianzen und Kovarianzen $\sigma_{jj'}$ impliziert, ist wie im Fall (a):

$$\Sigma_{n.} = \sigma^2_{N,t} \cdot (I_c - 1/c J_c) \quad .$$

In Willmes & Pyhel (1981) war noch eine sog. 'vollständige' Testprozedur zur detaillierten Analyse der (einfachen) Haupteffekte mit anschließenden aposteriori-Vergleichen vorgeschlagen worden. Die dort angegebene Strategie orientierte sich lediglich an der Kontrolle des globalen bzw. multiplen α-Niveaus. Im Sinn der Arbeiten von Bredenkamp (1980) und Hager & Westermann (1983 a, b, c) muß je nach Art der zu prüfenden psychologischen Hypothesen dieses Vorgehen modifiziert werden. Darauf wird in Kapitel 7.4 näher eingegangen.

(xii) Der zweifaktorielle vollständige Blockplan (RBF-rc, Stein-Zaar, 1983)

In einem zweifaktoriellen Blockplan (vgl. Kirk, 1968, Kap. 7.16) werden für jede von n Beobachtungseinheiten oder einen Block von homogenen Einheiten alle rc möglichen Kombinationen zweier Faktoren A und B mit r bzw. c Stufen angewendet. In Puri & Sen (1971, Kap. 7.7) wird die von Mehra & Sen (1969) entwickelte Klasse von alignierten Rangtests für das Wechselwirkungsproblem dargestellt. Der im folgenden hergeleitete Invarianz-Permutationstest besitzt bis auf die abschließende Rangtransformation große Ähnlichkeit mit obigem Test. Das Modell lautet:

$$Y_{hij} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\pi_h+\varepsilon_{hij} \qquad \begin{matrix} h=1,..,n \text{ Blöcke,} \\ i=1,..,r \text{ Stufen Faktor A,} \\ j=1,..,c \text{ Stufen Faktor B,} \\ N=nrc \end{matrix}$$

(3.69)

mit $\alpha_{\cdot}=\beta_{\cdot}=(\alpha\beta)_{i\cdot}=(\alpha\beta)_{\cdot j}=\pi_{\cdot}=0$ und unabhängigen

$\varepsilon_h=(\varepsilon_{h11},\ldots,\varepsilon_{hrc})'$,h=1,..,n, die verteilt sind nach symmetrischem $F_h(x)$.

Insgesamt gilt für die Verteilung F(x) von $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)'$,$x\in R^N$:

$$F(x) = F(\pi x) \quad \text{für alle } \pi\in \overset{n}{\underset{h=1}{\times}} S_{rc} \ .$$

Nimmt man identische F_h an, gilt $\pi\in S_{rc}\wr S_n$.

Getestet werden soll die Hypothese 'Keine Wechselwirkung':

$$H: (\alpha\beta)_{11} = \ldots = (\alpha\beta)_{rc} = 0$$

(3.70) gegen

$$A: (\alpha\beta)_{ij}\neq(\alpha\beta)_{i'j'} \text{ für ein } (i,j)\neq(i',j') \ .$$

In Analogie zu (3.4) hat man:

$$U = \{\theta\in R^N \mid \theta_{hij}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\pi_h;\ \mu,\alpha_i,\beta_j,(\alpha\beta)_{ij},\pi_h\in R\}$$

und

$$U^H = \{\theta\in U \mid \theta_{hij}=\mu+\alpha_i+\beta_j+\pi_h\} \ .$$

Man überlegt leicht, daß mit obigen Reparametrisierungsbedingungen gilt:

$$\theta\in U^H \iff \theta_{hij}-\bar\theta_{hi\cdot}-\bar\theta_{h\cdot j}+\bar\theta_{h\cdot\cdot} = 0 \quad \text{für alle } h,i,j$$

$$\iff (I_{nrc}-I_{nr}\otimes 1/cJ_c-I_n\otimes 1/rJ_r\otimes I_c+I_n\otimes 1/rcJ_{rc})\theta = 0$$

$$\iff K\theta = 0 \quad \text{mit } K=I_n\otimes(I_r-1/rJ_r)\otimes(I_c-1/cJ_c) \ .$$

Damit hat man als transformierte Zufallsvariablen:

$$(KY)_{hij} =: T_{hij} = Y_{hij}-\bar Y_{hi\cdot}-\bar Y_{h\cdot j}+\bar Y_{h\cdot\cdot} \ .$$

Hier sei im Vorgriff auf Abschnitt 3.2 darauf hingewiesen, daß T_{hij} eine Invariante aber keine Maximalvariante ist. Es gilt: $S_N(K)=S_r\otimes S_c\wr S_n$ aufgrund der Regeln in Pyhel (1978a, S. 11).

Für $\Omega_N^*=S_{rc}\wr S_n$ gilt $\Omega_N=S_N(K)$. Die Menge Ω_N der zulässigen Permutationen besteht aus Vertauschungen der ganzen Blöcke und zeilen- und spaltenweisen Vertauschungen innerhalb der Blöcke unabhängig von Block zu Block; Ω_N enthält $(r!c!)^n n!$

Elemente.

Wegen $\bar{\theta}_{.ij}-\bar{\theta}_{.i.}-\bar{\theta}_{..j}+\bar{\theta}_{...} = (\alpha\beta)_{ij}$ hat man:

$$K_A = 1/nJ_n \otimes ((I_r - 1/rJ_r) \otimes (I_c - 1/cJ_c))$$

und damit als Teststatistik:

$$(3.71) \quad S(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n\bar{T}^2_{.ij} \quad \text{oder äquivalent} \quad S^*(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{T}^2_{.ij} \,,$$

$$\text{denn } \bar{T}_{.i.} = \bar{T}_{..j} = \bar{T}_{...} = 0 \quad .$$

Die zulässigen Vertauschungen der ganzen Blöcke lassen die $\bar{T}_{.ij}$ und damit S(T) unverändert. Deshalb müssen zur Durchführung des Permutationstests pro Block jeweils unabhängig von Block zu Block nur die r!c! Zeilen- und Spaltenvertauschungen ausgeführt werden. Aus demselben Grund kann auch $Q^*_N = \overset{n}{\underset{h=1}{\times}} S_{rc}$-Invarianz der Fehler und damit $S_N(K) = \overset{n}{\underset{h=1}{\times}} (S_r \otimes S_c)$ betrachtet werden.

Mit denselben Methoden wie für den Split-Plot Plan kann für den RBF-rc Plan mit dem Permutationsmaß $P_N = (r!c!)^{-n}$ wieder eine quadratische Form in den $\bar{T}_{.ij}$ als Teststatistik berechnet werden. Ersetzt man in Puri und Sen (1971, S. 333) die transformierten Beobachtungen nicht durch Ränge, findet man mit:

$$\sigma^2_{N,t} := 1/(n(r-1)(c-1)) \sum_{h,i,j} t^2_{hij} \quad \text{und} \quad T_N := (\bar{T}_{.11}, \ldots, \bar{T}_{.rc})' \;:$$

$$E_{P_N}(T_N) = 0$$

und

$$V_N := \sigma^2_{N,t}/n \cdot (I_r - 1/rJ_r) \otimes (I_c - 1/cJ_c) \quad .$$

Eine verallgemeinerte Inverse von V_N ist $V_N^- = n/\sigma^2_{N,t} I_{rc}$ und damit:

$$Q(T_N) = n/\sigma^2_{N,t} \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{T}^2_{.ij} = S(T)/\sigma^2_{N,t} \quad .$$

Wie in Puri und Sen (1971, S. 331 unten) erwähnt, ist eine Erweiterung auf m>1 Replikationen je Zelle des Versuchsplans möglich. Das Modell lautet dann:

(3.72) $Y_{hijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \pi_h + \varepsilon_{hijk} \quad k=1,..,m>1$

und für die Verteilung der Fehler gelte:

$$F(x) = F(\pi x) \quad \text{für alle } \pi \in S_{rcm} \sim S_n \ .$$

Zum Test der Hypothese (3.70) findet man sofort:

$$\theta \in U^H \iff \theta_{hijk} - \bar{\theta}_{hi..} - \bar{\theta}_{h.j.} + \bar{\theta}_{h...} = 0$$

und

$$K^* = K \otimes 1/m J_m \ .$$

Weiterhin gilt: $Q_{Nm} = S_{Nm}(K^*) = S_m \sim S_N(K)$

und damit:

(3.73) $$S^*(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n \bar{T}^2_{.ij.} \ .$$

Die zulässigen Vertauschungen der T_{hijk} innerhalb der Zellen (i,j) je Block lassen $\bar{T}_{.ij.}$ und damit $S^*(T)$ invariant. Die weitere Analyse der (einfachen) Haupteffekte im Anschluß an den Test auf Wechselwirkungen wird in Abschnitt 7.4 erläutert.

(xiii) 3-faktorielle Versuchspläne - Test auf Tripel-interaktion

(CRF-rcs, RBF-rcs, SPF-rs.c; Stein-Zaar, 1983)

Die zuvor behandelten Lateinischen Quadrate und zweifach- bzw. partiell-hierarchischen Pläne enthielten auch drei Faktoren. Dort interessierten jedoch nur Tests auf bestimmte Haupteffekte bzw. Wechselwirkungen zwischen zwei Faktoren. Im folgenden soll immer der Test auf das Vorliegen der höchsten Wechselwirkung zwischen den drei Faktoren von Interesse sein. Obwohl es wenige psychologische Hypothesen

geben wird, die das Auftreten oder gar den Typ einer dreifachen Interaktion (Tripel-Interaktion, Interaktion 2. Ordnung) postulieren, vgl. z.B. Henning& Muthig, 1979), ist es dennoch wichtig, ihr Auftreten feststellen zu können. Je nach Art der dreifachen Interaktion ist bei Verallgemeinerung der Begriffe semi-disordinale und disordinale Interaktion (vgl,. Bredenkamp, 1980; Winer, 1971, Kap. 5.11) eine Interpretation von zweifachen Wechselwirkungen und/oder Haupteffekten nicht möglich. Dann muß zum Testen einfacher Interaktionen und einfacher (einfacher) Haupteffekte übergegangen werden (vgl. z.B. Kirk, 1968; Winer, 1971; Bortz, 1977; Henning & Muthig, 1979).

Für den dreifaktoriellen vollständigen Zufallsplan (CRF-rcs) ist kein nicht-trivialer Invarianz-Permutationstest auf dreifache Wechselwirkung angebbar, da schon im zweifaktoriellen Fall kein Test möglich war (vgl. Stein-Zaar, 1983). Andererseits ist für den dreifaktoriellen vollständigen Blockplan (RBF-rcs) direkt in Verallgemeinerung der Ergebnisse von (xii) ein Test auf dreifache Wechselwirkung möglich. Diese Behauptung findet man schon bei Puri & Sen (1971, S. 331 unten) für die Konstruktion eines entsprechenden Rangtests. Das lineare Modell lautet:

(3.74) $Y_{hijk}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk}+\pi_h+\varepsilon_{hijk}$

$h= 1,..,n$ Blöcke,
$i= 1,..,r$ Stufen Faktor A,
$j= 1,..,c$ Stufen Faktor B,
$k= 1,..,s$ Stufen Faktor C,

mit den üblichen Reparametrisierungsbedingungen und den ε_h unabhängig verteilt nach symmetrischen $F_h(x)=F_h(\pi x)$ für alle $\pi\in S_{rcs}$.

Insgesamt gilt für die Verteilung von ε, daß sie invariant ist gegenüber allen $\pi \in S_{rcs} \wr S_n$ bei identischen F_h; ansonsten ist $Q^*_n = \times_{h=1} S_{rcs}$.

Getestet werden soll $H: (\alpha\beta\gamma)_{111} = \ldots = (\alpha\beta\gamma)_{rcs} = 0$,

und man hat:

$$(3.75)\quad \theta \in U^H \iff \theta_{hijk} - \bar{\theta}_{hij.} - \bar{\theta}_{hi.k} - \bar{\theta}_{h.jk} + \bar{\theta}_{hi..} + \bar{\theta}_{h.j.} + \bar{\theta}_{h..k} - \bar{\theta}_{h...} = 0$$
$$\iff K\theta = 0 \quad \text{mit } K = I_n \otimes (I_r - 1/rJ_r) \otimes (I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s) .$$

Die transformierten Zufallsvariablen sind:

$$(KY)_{hijk} =: T_{hijk} = Y_{hijk} - \bar{Y}_{hij.} - \bar{Y}_{hi.k} - \bar{Y}_{h.jk} + \bar{Y}_{hi..} + \bar{Y}_{h.j.} + \bar{Y}_{h..k} - \bar{Y}_{h...}$$

Die zulässigen Permutationen $Q_N = S_N(K) = (S_r \otimes S_c \otimes S_s) \wr S_n$ bestehen aus Vertauschungen der ganzen Blöcke, die die Teststatistik invariant lassen, und zeilen-, spalten- und 'Scheiben'-weisen Vertauschungen innerhalb der Blöcke unabhängig von Block zu Block. Q_n enthält $(r!c!s!)^n \cdot n!$ Elemente, und eine geeignete Teststatistik ist:

$$(3.76)\quad S(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{s} n\bar{T}^2_{.ijk} \quad .$$

Eine Erweiterung auf m>1 Replikationen je Zelle je Block ist wieder analog zu (3.73) möglich.

Aus der Struktur von K in (3.75) ist ersichtlich, daß eine Verallgemeinerung auf mehr als 3 Faktoren mit Meßwiederholungen sofort möglich ist. Es schließen sich 'rechts' durch Kronecker-Produkte verknüpfte Matrizen derselben Struktur $(I_\nu - 1/\nu J_\nu)$ an und in Q_n werden ebenfalls S_ν 'rechts' angehängt:

$$Q_N = (S_r \otimes S_c \otimes S_s \otimes S_\nu \otimes \ldots) \wr S_n \quad .$$

Eine geeignete Teststatistik ergibt sich stets durch Bildung der Summe aller quadrierten 'Zellensummen' der transformierten Beobachtungen über alle n Blöcke hinweg.

Für den Split-Plot-Plot Plan (SPF-rc.s) hat Stein-Zaar (1983) gezeigt, daß ein Invarianz-Permutationstest auf die Tripelinteraktion nicht existiert. Die Bestimmung der zulässigen Permutationen $S_N(K)$, unter denen die Teststatistik dann jedoch invariant ist, läßt sich einfacher ausführen als bei ihr angegeben. Deshalb soll das Vorgehen auch für diesen Plan kurz erläutert werden.

Den Split-Plot-Plot Plan mit o.B.d.A. identischer Anzahl n>1 von Blöcken je Plotstufen-Kombination kann man sich wie in Kirk (1968, S. 284) veranschaulichen.

Die Modellgleichung lautet:

$$(3.77)\quad Y_{ijh(i,j)k} = \mu+\alpha_i+\beta_j+\gamma_k+\pi_{h(i,j)}+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+ \\ +(\alpha\beta\gamma)_{ijk}+(\pi\gamma)_{h(i,j)k}+\varepsilon_{ijh(i,j)k}$$

$i= 1,..,r$ Plot-Faktor A,
$j= 1,..,c$ Plot-Faktor B,
$k= 1,..,s$ Split-Faktor C,
$h(i,j)= 1,..,n$ Blöcke je Plot-Kombination (i,j)

mit den üblichen Reparametrisierungsbedingungen und S_S-invarianten Fehlern $\varepsilon_{ijh(i,j)}$, die unabhängig und identisch verteilt seien. Insgesamt bedeutet das für die Verteilung F von ε: $F(x)=F(\pi x)$ für alle $\pi\in Q_N^*=S_s\wr S_{rcn}$.

Getestet werden soll:

$$H:\quad (\alpha\beta\gamma)_{ijk} = 0 \qquad \text{für alle } i,j,k\ .$$

Mit den üblichen Definitionen von θ und U, U^H gilt:

$$\theta\in U^H \iff \theta_{ijh(i,j)k}-\bar{\theta}_{ijh(i,j).}-\bar{\theta}_{i..k}-\bar{\theta}_{.j.k}+\bar{\theta}_{i...}+\bar{\theta}_{.j..}+\bar{\theta}_{...k}- \\ -\bar{\theta}_{....} = 0\ .$$

In Matrixschreibweise bedeutet das: $K\theta=0$ mit

$$K=I_{rcns}-I_{rcn}\otimes 1/sJ_s-I_r\otimes 1/cnJ_{cn}\otimes I_s-1/rJ_r\otimes I_c\otimes 1/nJ_n\otimes I_s+ \\ +I_r\otimes 1/cnsJ_{cns}+1/rJ_r\otimes I_c\otimes 1/nsJ_{ns}+1/rcnJ_{rcn}\otimes I_s-1/rcnsJ_{rcns}$$

(3.78) $=(I_{rcn}-I_r\otimes 1/cnJ_{cn}-1/rJ_r\otimes I_c\otimes 1/nJ_n+1/rcnJ_{rcn})\otimes(I_s-1/sJ_s)$.

Der Ausdruck in der ersten Klammer ist identisch mit der Transformationsmatrix in (3.31) für den CRF-rc Plan. Dieser hatte die Invarianzstruktur $S_n\wr(S_r\otimes S_c)$. Insgesamt ist $S_n(K)$ für K aus (3.78) dann $(S_n\wr S_r\otimes S_c)\otimes S_s$.
Damit sind wie beim CRF-rc (vgl. (3.32)!) neben Permutationen der Blöcke je Faktorstufenkombination A_iB_j und von Split-Faktorstufen über alle Blöcke hinweg nur dieselben Vertauschungen der Stufen je eines Plot-Faktors auf allen Stufen des anderen Plot-Faktors zulässig. Diese bewirken in der geeigneten Teststatistik

$$S(T) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n\bar{T}^2_{ij.k}$$

jedoch nur Umindizierungen.

Mit diesem Ergebnis ist auch für alle höherfaktoriellen Split-Plot Pläne mit mindestens 2 gekreuzten Plot-Faktoren ein Invarianz-Permutationstest auf die höchste Wechselwirkung mit den dargestellten Methoden nicht möglich.

Der Versuchsplan für den Split-Split-Plot Plan (SPF-r.cs) mit pro Plot-Stufe $n_i>1$ Blöcken läßt sich wie in Kirk (1968, S. 299) darstellen.
Ein solcher Versuchsplan mit faktorieller Struktur für die wiederholten Messungen ist in verschiedenen Bereichen der Psychologie häufig zu finden, in denen Versuchspersonen ein 'knappes Gut' sind. Bock (1966, S. 481) nennt z.B. Lernpsychologie, Wahrnehmungsforschung und Psycholinguistik.
Die Modellgleichung lautet:

(3.79) $y_{ih(i)jk} = \mu + \alpha_i + \beta_j + \gamma_k + \pi_{h(i)} + (\pi\beta)_{h(i)j} + (\pi\gamma)_{h(i)k} + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ih(i)jk}$

$i = 1,..,r$ Plot-Faktor A,
$j = 1,..,c$ Split-Faktor B,
$k = 1,..,s$ Split-Faktor C,
$h(i) = 1,..,n$ Blöcke je Plot-Stufe A_i

mit den üblichen Reparametrisierungsbedingungen und S_{cs}-invarianten Fehlern $\varepsilon_{ih(i)}$, die identisch verteilt seien. Insgesamt sei die Verteilung von ε $\Omega_N^* = S_{cs} \wr S_{n.}$-invariant.

Zu testen ist wieder die Hypothese:

$H: \quad (\alpha\beta\gamma)_{ijk} = 0 \qquad$ für alle i,j,k .

Mit den üblichen Definitionen von θ, U, U^H hat man:

$$\theta \in U^H \iff \theta_{ih(i)jk} - \bar{\theta}_{ih(i)j.} - \bar{\theta}_{ih(i).k} - \bar{\theta}_{..jk} + \bar{\theta}_{ih(i)..} + \bar{\theta}_{..j.} + \bar{\theta}_{...k} - \bar{\theta}_{...} = 0$$

$$\iff K\theta = 0 \quad \text{mit}$$

$$K = I_{n.cs} - I_{n.c} \otimes 1/sJ_s - I_{n.} \otimes 1/cJ_c \otimes I_s - 1/n.J_{n.} \otimes I_{cs} +$$

$$+ I_{n.} \otimes 1/csJ_{cs} + 1/n.J_{n.} \otimes I_c \otimes 1/sJ_s + 1/n.cJ_{n.c} \otimes I_s -$$

$$- 1/n.csJ_{n.cs}$$

$$= (I_{n.} - 1/n.J_{n.}) \otimes (I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s) \quad .$$

Damit ist in direkter Verallgemeinerung der Ergebnisse für den Split-Plot Plan:

(3.80) $\Omega_N = S_N(K) \cap \Omega_N^* = S_{n.} \otimes (S_c \otimes S_s) \cap S_{cs} \wr S_{n.} = S_{n.} \otimes (S_c \otimes S_s)$.

Es dürfen also alle n. cs-dimensionalen transformierten Blöcke untereinander über alle Stufen des Plot-Faktors hinweg vertauscht werden. Zusätzlich können dieselben Stufen jeweils jedes der beiden Split-Faktoren über <u>alle</u> n. Blöcke hinweg permutiert werden. Ω_n enthält n.!c!s! Elemente.

Zur Konstruktion einer Teststatistik überlegt man:

$$\bar{\theta}_{i.jk} - \bar{\theta}_{i.j.} - \bar{\theta}_{i..k} - \bar{\theta}_{..jk} + \bar{\theta}_{i...} + \bar{\theta}_{..j.} + \bar{\theta}_{...k} - \bar{\theta}_{....} = (\alpha\beta\gamma)_{ijk}$$

und damit

$$K_A = \bigotimes_{i=1}^{r} (1/n_i J_{n_i} \otimes ((I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s))) - 1/n.J_{n.} \otimes ((I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s)) \quad .$$

Angewendet auf T ergibt sich:

$$K_A T = (\bar{T}_{1.11},\ldots,\bar{T}_{1.11},\ldots,\bar{T}_{r.cs},\ldots,\bar{T}_{r.cs})'$$

n_1 - mal $\qquad\qquad$ n_r - mal

und als Teststatistik:

$$(3.81) \qquad S(T) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n_i \bar{T}^2_{i.jk} \quad .$$

S(T) ist invariant unter den zulässigen Vertauschungen aus $S_c \otimes S_s$, die - weil auf alle n. Blöcke angewendet - nur eine Umindizierung unter den Indizes j und k bewirken. S(T) ist dagegen nicht invariant unter den n.! möglichen Permutationen der ganzen cs-dimensionalen transformierten Blöcke. Da für (3.81) in

$$\bar{T}_{i.jk} = \bar{Y}_{i.jk} - (\bar{Y}_{i.j.} - \bar{Y}_{i..k} - \bar{Y}_{i...}) - (\bar{Y}_{..jk} - \bar{Y}_{..j.} - \bar{Y}_{...k} + \bar{Y}_{....})$$

die letzten 4 Terme invariant sind unter $\mathcal{G}_N$ aus (3.80), genügt es, anstelle von (3.81)

$$(3.82) \qquad S(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n_i (\bar{Y}_{i.jk} - \bar{Y}_{i.j.} - \bar{Y}_{i..k} + \bar{Y}_{i...})^2$$

zu betrachten.

In dieser Form sieht man auch den Zusammenhang zum RBF-cs Plan. In (3.82) hat man die Summe von r mit den jeweiligen Stichprobenumfängen gewichteten RBF-cs Statistiken (vgl. (xii)!). Die S(Y) nicht invariant lassenden $n_\cdot ! / \prod_{i=1}^{r} n_i !$ Blockpermutationen über verschiedene Split-Stufen hinweg entsprechen denen des CR-r Planes. Stein-Zaar (1983) war aus folgendem Grund zu einem negativen Ergebnis bei der Konstruktion eines Wechselwirkungs-Tests für diesen Plan gekommen: Nimmt man die einfache Wechselwirkung zwischen den Blöcken und jedem der beiden Split-Effekte nicht in die Modellgleichung mit auf, erscheint der Term $2\bar{Y}_{i...}$ in der

Gleichung für $(KY)_{ih(i)jk}$. Dieser Ausdruck verhindert bei der Bildung von $S_N(K)$ eine Vertauschung der Blocke uber die Plotstufen hinweg, sodaß sowohl (3.81) wie auch die bei Stein-Zaar gewählte Teststatistik invariant bleiben unter den dann noch zulässigen Permutationen aus $(\bigtimes_{i=1}^{r} S_{n_i})\otimes S_c\otimes S_s$. Es sei wieder darauf hingewiesen, daß $T_{ih(i)jk}$ eine Invariante, aber keine Maximalinvariante darstellt. Eine Verallgemeinerung auf mehr als zwei Split-Faktoren ist wieder direkt möglich. Es ist nur notwendig, die Blockeffekte und ihre einfachen und mehrfachen Wechselwirkungen mit den Split-Faktoren in die Modellgleichung mit aufzunehmen (vgl. z.B. Kirk, 1968, S. 308 für den SPF-r.csu). In der Transformationsmatrix K werden lediglich 'rechts' in der Klammer mit Kroneckerprodukt ein oder mehrere Ausdrücke der Form $(I_\nu - 1/\nu J_\nu)$ angeschlossen. Die für die Existenz eines nichttrivialen Invarianz-Permutationstests erforderliche Vertauschbarkeit aller transformierten Blocke wird davon nicht beeinflußt, sodaß $Q_N = S_{n_\cdot}\otimes S_c\otimes S_s\otimes \ldots$. Eine geeignete Teststatistik ist dann wieder $\sum_{i,j,k,l,\ldots} n_i \bar{T}^2_{i.jkl\ldots}$.

Anstelle der Hypothese 'keine Tripelinteraktion' kann wieder in Analogie zu (3.63) eine verallgemeinerte Profilparallelitätshypothese getestet werden. Untersucht werden soll, ob für die verschiedenen experimentellen Bedingungen des (Plot-) Faktors A die Relationen bezüglich der Split-Stufen-Kombinationen identisch sind. Es könnte z.B. überprüft werden, ob bei zufällig den Bedingungen 'Lärmbelästigung ja/nein' (Plot Faktor A) zugewiesenen Vpn morgens/abends (Split Faktor B) vor/mitten in/nach (Split-Faktor C) einer 2-stündigen Arbeitsphase die Relation der beiden (mittleren) Leistungsprofile (morgens/abends) in

einem Konzentrationstest zueinander identisch ist oder nicht. Niveauunterschiede sind dagegen für diese Frage zulässig. Die Situation für verallgemeinerte Profilparallelität ist in Abb. 3.1 schematisch dargestellt (vgl. Winer, 1971, S. 182f).

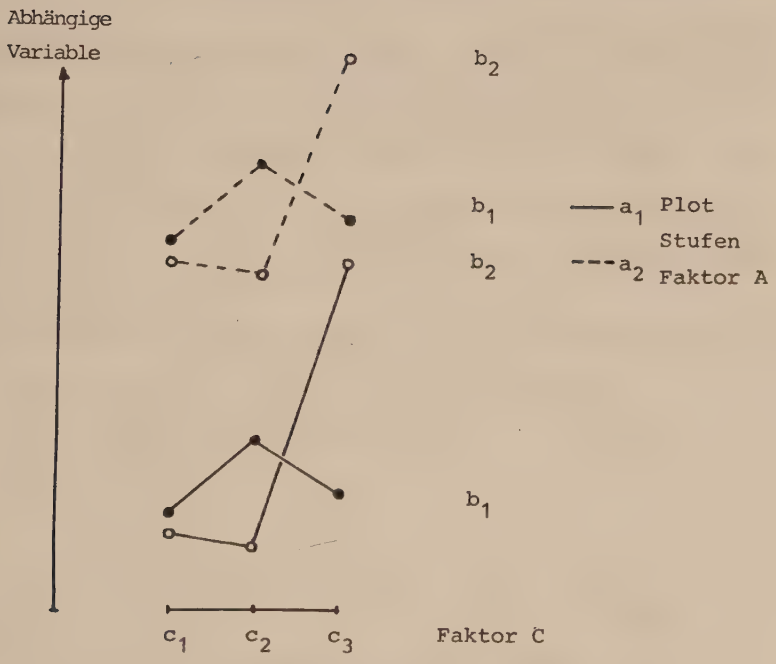


Abb. 3.1. Schematische Darstellung einer fehlenden Tripelinteraktion bzw. Darstellung der verallgemeinerten Profilparallelität im Split-Split-Plot-Plan; angegeben sind die hypothetischen Zellmittel.

Als Modell hat man analog zum Fall (b) in (xi):

$$Y_{ih(i)jk} = \mu_{ih(i)jk} + \epsilon_{ih(i)jk} \qquad \begin{matrix} i,h(i),j,k \text{ wie in (3.79);} \\ n=\sum n_i;\ N=n_. cs \end{matrix}$$

mit den $n_.$ verallgemeinerten cs-dimensionalen Profilen

$$\mu_{ih(i)jk} = (\mu_{ih(i)11},\ldots,\mu_{ih(i)cs})' \quad .$$

Wiederum wird für die Fehler-Blöcke nur angenommen, daß sie über alle $n_.$ Profile vertauschbar sind, d.h. $F(x) = F(\pi x)$ für alle $\pi \in \tilde{\tilde{S}}_{n_.} := S_{n_.} \otimes \{1_{cs}\}$. $\tilde{\tilde{S}}_{n_.}$ enthält $n_.!$ Vektor-Permutationen. Niveauunterschiede jedoch keine Variabilitätsunterschiede zwischen den Blöcken sind zugelassen sowie unterschiedliche Variabilität für die Split-Stufen Kombinationen.

Die Hypothese der verallgemeinerten Profilparallelität für alle $n_.$ Profile gegen die Alternative, daß innerhalb jeder Plotstufe für die n_i-Profile zwar Parallelität, aber für mindestens zwei Plotstufen i,i' keine Parallelität vorliegt, läßt sich so formulieren:

(3.83) H: Es gibt $(\Delta_{11},\ldots,\Delta_{c-1,s-1})' \in R^{(c-1)(s-1)}$ mit

$$\mu_{ih(i)jk} - \mu_{ih(i)j,k+1} - (\mu_{ih(i)j+1,k} - \mu_{ih(i)j+1,k+1}) = \Delta_{jk}$$

für alle $j=1,..,c-1$
$k=1,..,s-1$
$h(i)=1,..,n_i$
$i=1,..,r$

gegen

A: $\mu_{ih(i)jk} - \mu_{ih(i)j,k+1} - (\mu_{ih(i)j+1,k} - \mu_{ih(i)j+1,k+1}) = \Delta_{ijk}$
und $\Delta_{ijk} \neq \Delta_{i'jk}$ für ein $i \neq i' \in \{1,..,r\}$.

Für diese Hypothese gibt es wie in (3.63) eine äquivalente Darstellung als:

(3.84) H: $\mu_{ih(i)jk} - \bar{\mu}_{ih(i)j.} - \bar{\mu}_{ih(i).k} + \bar{\mu}_{ih(i)..} - (\bar{\mu}_{..jk} - \bar{\mu}_{..j.} - \bar{\mu}_{...k} + \bar{\mu}_{...}) = 0$

gegen

A: $\mu_{ih(i)jk} - \bar{\mu}_{ih(i)j.} - \bar{\mu}_{ih(i).k} + \bar{\mu}_{ih(i)..} - (\bar{\mu}_{..jk} - \bar{\mu}_{..j.} - \bar{\mu}_{...k} + \bar{\mu}_{....}) = \nu_{ijk} \neq 0$
für ein Tripel (i,j,k) und alle $h(i)=1,..,n_i$.

Zum Beweis überlegt man Folgendes:

Durch Ausnutzen der Beziehung für Δ_{jk} in (3.83) und mit

$\mu_{ih(i)jk} - \mu_{ih(i)j,k+1} =: \eta_k^j$ sowie $\mu_{ih(i)jk} - \mu_{ih(i)j+1,k} =: \xi_j^k$
zeigt man nach umfangreichen algebraischen Umformungen,

falls H als gültig angenommen wird:

$\bar{\mu}_{ih(i)..} = 1/cs \sum_{j'} \sum_{k'} \mu_{ih(i)j'k'} = \mu_{ih(i)jk} + \Delta_{jk}^{**}$ für jedes $j=1,..,c-1$, $k=1,..,s-1$

und alle (i,h(i)) mit Δ_{jk}^{**} abhängig von η_k^j und ξ_j^k .

Damit gilt auch:

$\bar{\mu}_{....} = 1/n \cdot \sum_i \sum_{h(i)} \bar{\mu}_{ih(i)..} = \bar{\mu}_{..jk} + \Delta_{jk}^{**}$.

Außerdem hat man in Verallgemeinerung von Pyhel (1980a):

$\bar{\mu}_{ih(i).k} = 1/c \sum_{j'=1}^{c} \mu_{ih(i)j'k} =: \mu_{ih(i)jk} + \Delta_j^{*k}$ für jedes $j=1,..,c-1$, $k=1,..,s$.

$\bar{\mu}_{ih(i)j.} = 1/s \sum_{k'=1}^{s} \mu_{ih(i)jk'} =: \mu_{ih(i)jk} + \Delta_k^{*j}$ für jedes $k=1,..,s-1$, $j=1,..,c$.

Damit gilt:

$\bar{\mu}_{...k} = 1/n \cdot \sum_i \sum_{h(i)} \bar{\mu}_{ih(i).k} = \bar{\mu}_{..jk} + \Delta_j^{*k}$; $\bar{\mu}_{..j.} = \bar{\mu}_{..jk} + \Delta_k^{*j}$.

Insgesamt also:

$$\begin{aligned} & \mu_{ih(i)jk} - \bar{\mu}_{ih(i)j.} - \bar{\mu}_{ih(i).k} + \bar{\mu}_{ih(i)..} - (\bar{\mu}_{..jk} - \bar{\mu}_{..j.} - \bar{\mu}_{...k} + \bar{\mu}_{....}) \\ = \; & \mu_{ih(i)jk} - \mu_{ih(i)jk} - \Delta_k^{*j} - \mu_{ih(i)jk} - \Delta_j^{*k} + \mu_{ih(i)jk} + \Delta_{jk}^{**} - \bar{\mu}_{..jk} + \\ & + \bar{\mu}_{..jk} + \Delta_k^{*j} + \bar{\mu}_{..jk} + \Delta_j^{*k} - \bar{\mu}_{..jk} - \Delta_{jk}^{**} \\ = \; & 0 \end{aligned}$$

Andererseits hat man, wenn (3.84) gilt:

$$\mu_{ih(i)jk} - \mu_{ih(i)j+1,k} - \mu_{ih(i)j,k+1} + \mu_{ih(i)j+1,k+1}$$

$$= -(\bar{\mu}_{..jk} - \bar{\mu}_{..j+1,k} - \bar{\mu}_{..j,k+1} + \bar{\mu}_{..j+1,k+1}) =: \Delta_{jk} \quad .$$

In Matrixschreibweise bedeutet (3.81):

$$K\mu = 0 \quad \text{mit} \quad K = (I_{n_.} - 1/n_. J_{n_.}) \otimes ((I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s)) \quad .$$

Damit hat man $Q_N = S_N(K) \cap \tilde{S}_{n_.} = S_{n_.} \otimes \{1_{rcs}\}$ und dieselbe Teststatistik wie in (3.82) mit $n_.! / \prod_{i=1}^{r} n_i!$ möglichen verschiedenen Werten unter den zulässigen $n_.!$ Permutationen der ganzen Blöcke. Wie für den Split-Plot Plan bedeutet die restriktive Annahme an die Fehler-Blöcke keine Einschränkung der Menge zulässiger Permutationen, die die Teststatistik nicht invariant lassen.

Auch die Konstruktion einer quadratischen Form als Teststatistik erfolgt für den Test auf Tripelinteraktion wie auf verallgemeinerte Profilparallelität analog zum Split-Plot Plan.

Fall (a): Bei gegebener Realisation t haben, wenn die $t_{ih(i)jk}$ wie im Versuchsplan angeordnet werden, alle Zeilenvertauschungen und alle Permutationen für die Stufen jedes der beiden Split-Faktoren dieselbe Wahrscheinlichkeit $P_N = 1/n_.!c!s!$. Dann gilt für die ein- und zweidimensionalen Randverteilungen von T bei gegebenem t:

$$P_N(T_{ih(i)jk} = x) = 1/n_.cs$$

$$P_N(T_{ih(i)jk} = x,\ T_{i'h'(i')j'k'} = x') = \begin{cases} 1/n_.cs(c-1)(s-1) & (i,h) = (i',h') \\ & j \neq j',\ k \neq k' \\ 1/n_.cs(s-1) & (i,h) = (i',h') \\ & j = j',\ k \neq k' \\ 1/n_.cs(c-1) & (i,h) = (i',h') \\ & j \neq j',\ k = k' \\ 1/n_.(n_.-1)cs & (i,h) \neq (i',h');\ (j,k) = \\ 1/n_.(n_.-1)cs(c-1)(s-1) & " \quad (j,k) \neq \\ 1/n_.(n_.-1)cs(s-1) & " \quad j = j'; \\ 1/n_.(n_.-1)cs(c-1) & " \quad j \neq j'; \\ 0 & \text{sonst} \end{cases}$$

Nach umfangreichen algebraischen Umformungen in Analogie zum Test auf Wechselwirkungen im Split-Plot Plan erhält man für $T_N := (\bar{T}_{1.11}, \ldots, \bar{T}_{r.cs})'$:

$$V_N := E_{P_N}(T_N \cdot T_N') = (\sigma^2_{N,t}/n_.) \cdot ((n_./n_i \delta_{ii'} - 1)) \otimes ((I_c - 1/cJ_c) \otimes (I_s - 1/sJ_s))$$

$$\text{mit } \sigma^2_{N,t} := 1/(n_. - 1)(c-1)(s-1) \sum_{i,h(i),j,k} t^2_{i,h(i),j,k}$$

und nach Bildung einer verallgemeinerten Inversen von V_N als quadratische Form

$$(3.85) \quad Q(T_N) = 1/\sigma^2_{N,t} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{s} n_i \bar{T}^2_{i.jk} = S(T)/\sigma^2_{N,t} \text{ wegen (3.81)} .$$

Asymptotisch besitzt $Q(T_N)$ eine χ^2-Verteilung mit df = $(r-1) \cdot (c-1)(s-1)$. Der Beweis erfolgt in Analogie zum Theorem (7.9) von Pyhel (1978b).

Fall (b): Alle Vertauschungen von Zeilen im Versuchsschema für die $t_{ih(i)jk}$ besitzen dieselbe Wahrscheinlichkeit 1/n.!. Damit gilt für die ein- und zweidimensionalen Randverteilungen:

$$P_{n_.}(T_{ih(i)jk} = x) = 1/n_.$$

$$P_{n_.}(T_{ih(i)jk} = x; T_{i'h'(i')j'k'} = x') = \begin{cases} 1/n_. & (i,h) = (i',h') \\ & (j,k) \neq (j',k') \\ 1/n_.(n_.-1) & (i,h) \neq (i',h') \\ & (j,k), (j',k') \text{ beliebig} \\ 0 & \text{sonst} \end{cases}$$

Wiederum berechnet man:

$$V_{n_.} = 1/n_.(n_./n_i \delta_{ii'} - 1) \otimes \Sigma_{n_.} \quad \text{mit} \quad \Sigma_{n_.} = ((\sigma_{jj'kk'}))_{cs \times cs}$$

und $$\sigma_{jj'kk'} := 1/(n_. - 1) \sum_{i=1}^{r} \sum_{h(i)=1}^{n_i} t_{ih(i)jk} \cdot t_{ih(i)j'k'}$$

und erhält als Teststatistik:

$$(3.86) \quad Q(T_N) = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{j'=1}^{c} \sum_{k=1}^{s} \sum_{k'=1}^{s} n_i \bar{T}_{i.jk} \, \sigma^{**}_{jj'kk'} \, \bar{T}_{i.j'k'}$$

mit $\sigma^{**}_{jj'kk'}$ als Element einer (verallgemeinerten) Inversen von $\Sigma_{n.}$. $Q(T_N)$ ist asymptotisch χ^2_f - verteilt mit $(r-1)\leq f\leq(r-1)(c-1)(s-1)$ Freiheitsgraden in Analogie zu Theorem (7.7) in Pyhel (1980a).

Fall (c): Ein Spezialfall von (b) soll noch näher erläutert werden. Falls es sich nicht um ein verallgemeinertes Profil handelt, sondern o.B.d.A. die Stufen des ersten Split-Faktors randomisiert den Beobachtungseinheiten zugewiesen werden können und nur der zweite 'Split-Faktor' ein Profil (z.B. verschiedene Zeitpunkte) darstellt, hat man gegenüber (3.86) vereinfachend:

$$(3.87)\qquad Q(T_N) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s}\sum_{k'=1}^{s} n_i \bar{T}_{i.jk}\,\sigma^{*}_{kk'}\,\bar{T}_{i.jk'}$$

mit $\sigma^{*}_{kk'}$ Element der sxs-dimensionalen (verallgemeinerten) Inversen von $((\sigma_{kk'}))_{s\times s}$ und

$$\sigma_{kk'} = (1/(n_.-1)(c-1))\sum_{i=1}^{r}\sum_{h(i)=1}^{n_i}\sum_{j=1}^{c} t_{ih(i)jk}\cdot t_{ih(i)jk'}\ .$$

Dieses $Q(T_N)$ ist asymptotisch χ^2_f - verteilt mit $(r-1)(c-1)\leq f\leq(r-1)(c-1)(s-1)$ Freiheitsgraden.

3.1.6 Zusammenfassung und Diskussion

Für viele in gängigen Lehrbüchern beschriebene ein- bis dreifaktorielle gekreuzte, hierarchische und partiell hierarchische Versuchspläne mit festen Effekten wurde die Existenz von Invarianz-Permutationstests untersucht. Da kein allgemeines konstruktives Verfahren bekannt ist, muß für die verschiedenen linearen Hypothesen über Teile des gesamten Parametervektors in den verschiedenen Versuchsplänen stets neu überprüft werden, ob die Konstruktion eines Invarianz-Permutationstests möglich ist. Der Ansatz von Pyhel, in einem 'verteilungsfreien' linearen Modell durch geeignete lineare Transformationen der Beobachtungen die 'störenden' Parameter zu eliminieren und die Permutationsinvarianz-eigenschaften der Verteilungsfunktion der transformierten Zufallsvariablen zur Testkonstruktion zu nutzen, wurde für alle betrachteten Pläne verwendet. Damit konnte eine zusammenfassende Darstellung bisher nur zum Zeil publizierter Ergebnisse für einzelne Versuchspläne gegeben und für einige lineare Hypothesen neue Invarianz-Permutationstests vorgeschlagen werden (CRH-r(c) s, SPF-r.cs).

Still & White (1981) schlagen ebenfalls eine Eliminierung störender Effekte vor. Sie übersehen jedoch, daß die Permutationsinvarianzeigenschaften der resultierenden Residuen so verändert sein können, daß wie im zweifaktoriellen vollständig gekreuzten Zufallsplan mit den dargestellten Methoden kein nicht-trivialer Invarianz-Permutationstest auf Wechselwirkung mehr angebbar ist (vgl. a.a.O., S. 248 oben). Generell gilt, daß in allen Versuchsplänen, die mindestens zwei gekreuzte Faktoren ohne Meßwiederholungen enthalten, ein Test auf die höchste Wechselwirkung nicht möglich ist.

Das war neben dem zweifaktoriellen Zufallsplan für den Split-Plot-Plot Plan und den partiell hierarchischen Plan CRH-r(c)s der Fall.

Für Versuchspläne mit Meßwiederholungen und höchstens einem Faktor ohne Meßwiederholungen war stets ein Invarianz-Permutationstest auf die höchste Wechselwirkung möglich. Der Grund ist, daß im wesentlichen die störenden Effekte immer (nur) pro Block eliminiert werden. Die zulassigen invarianten aber nicht maximalinvarianten Transformationen können deshalb, da die Blöcke unabhängig permutierbar sind, die zulässigen Permutationsgruppen nicht so sehr einschränken, daß nur noch eine geeignete Teststatistik invariant lassende Vertauschungen übrig bleiben.

Eine Sonderstellung nehmen die Tests auf (verallgemeinerte) Profilparallelität ein. In diesen Plänen gibt es einen oder zwei Faktoren mit Meßwiederholungen, für die keine randomisierte Zuweisung der Reihenfolge der Faktorstufen zu den Beobachtungseinheiten möglich (z.B. Faktor Zeitpunkte) oder von Interesse (z.B. Testprofil) ist. Die Profilparallelitätshypothesen lassen sich so darstellen, daß dieselbe Transformationsmatrix wie im Split-Plot bzw. Split-Split-Plot Versuchsplan resultiert. Die Teststatistik kann in beiden Fällen wie bei experimenteller Randomisierung der Faktorstufen gebildet werden, da Vertauschungen der Split-Stufe(n) diese invariant lassen. Für beide Profilparallelitätstests wurden alternativ zur ANOVA-analogen Teststatistik auch quadratische Formen als Teststatistiken hergeleitet. In diesen wird die mögliche Abhängigkeit zwischen den Profilstufen mitberücksichtigt.

Verglichen mit den ANOVA-Modellen sind in den hier betrachteten sog. verteilungsfreien Modellen lediglich die Voraussetzungen an die Verteilungsfunktion der Fehler abgeschwächt worden. Unabhängigkeit und Normalverteiltheit (mit i.a. identischen Varianzen) sind ersetzt worden durch bestimmte Permutationsinvarianzannahmen für die gemeinsame Verteilungsfunktion des gesamten Fehlervektors. Theoretisch ist diese Voraussetzung schwächer. Inwieweit aber z.B. im Split-Plot Plan Permutationsinvarianzen der Block-Fehlervektoren immer gerechtfertigt sind, sei hier nur kritisch angemerkt. Die lineare Modellstruktur mit linearen Hypothesen ist beibehalten worden. Allerdings zeigen neuere Arbeiten zu Rangtests in Multiparametermodellen, daß ein Abgehen von solchen (sog. Lokations-) Modellen in mehrfaktoriellen Plänen zu ganz anderen Hypothesen führt (Hilgers, 1979, 1981; sowie im Überblick Erdfelder & Bredenkamp, 1984).

3.2. Diskussion des Konstruktionsprinzips für Invarianz-Permutationstests

Eine einfachere Darstellung des Konstruktionsprinzips für Invarianz-Permutationstests kann man geben, wenn man eine etwas abstraktere Fassung des linearen Modells wählt. Diese schließt sich an Zyskind (1967), Witting & Nölle (1970) und Roebruck (1978, 1979, 1982a, 1983) sowie Roebruck & Willmes (1983) an. Die wichtigsten Ableitungsschritte aus Abschnitt 3.1 werden zur Übersicht nochmals wiederholt.

Wie zuvor ist ein verteilungsfreies lineares Modell gegeben durch

$$Y = \theta + \varepsilon \quad .$$

θ stamme aus einem linearen Teilraum U des R^n. Für die Verteilungsfunktion F des Fehlers sei nur angenommen, daß $F(x)=F(\pi x)$ mit π Element einer nicht notwendig echten Untergruppe $\mathfrak{Q}^*_n$ der symmetrischen Gruppe S_n.

Das Testproblem sei wieder

$$H: \theta \in U^H \quad \text{gegen} \quad A: \theta \notin U^H \quad .$$

Dieses Testproblem ist invariant unter der Gruppe von Verschiebungen

$$G = \{g \mid g(Y) = Y + \theta^*, \theta^* \in U^H\} \quad .$$

Bei Wahl (irgendeiner) G-invarianten Statistik $T: R^n \to R^n$, die den n-dimensionalen Beobachtungsvektor in einen anderen n-dimensionalen Zufallsvektor überführt, gilt unter H:

$$T(Y) = T(\varepsilon) \quad .$$

Eine geeignete Statistik ist die zuvor gewählte orthogonale Projektion in den zu U^H orthogonalen Raum $(U^H)^\perp$.

Gesucht ist nun eine Permutationsgruppe $\mathfrak{Q}_n$, für die gilt: Unter H ist die induzierte Verteilung der Statistik permutationsinvariant für alle $\pi \in \mathfrak{Q}_n$:

$$(F^T)^\pi = F^T \quad .$$

Eine hinreichende (aber nicht notwendige) Bedingung für diese Eigenschaft ist wie zuvor schon dargestellt:

Falls zu $\pi \in \mathfrak{Q}_n$ ein $\rho \in \mathfrak{Q}^*_n$ existiert mit $\pi T = T\rho$ gilt:
$(F^T)^\pi = (F^\rho)^T = F^T$ wegen der $\mathfrak{Q}^*_n$-Invarianz von F.

Wie in Abschnitt 3.1.3 schon erwähnt, hat Pyhel (1978a) gezeigt, daß für orthogonale Projektionen als Statistik T sogar gilt: $\pi T = T\pi$, d.h. es handelt sich um jeweils dieselbe Permutation.

Entscheidend für die Existenz eines Invarianz-Permutationstests ist demnach, daß auch der sog. Bildraum der Statistik T eine geeignete Permutationsstruktur charakteristiert durch $\mathcal{Q}_n$ aufweist.

Letzter Schritt ist die Wahl einer geeigneten reellwertigen Teststatistik S(T). Nimmt man dazu den Zähler der ANOVA-F-Statistiken, die im Normalverteilungsfall gewisse Optimalitätseigenschaften besitzen, gilt sogar S(T)=S(Y), wie in (3.16) dargestellt.

Man benötigt die Transformation T(Y) also 'nur' um festzustellen, ob in $\mathcal{Q}_n$ 'genügend' Permutationen enthalten sind, um einen nichttrivialen Permutationstest mit Permutationsmaß $1/|\mathcal{Q}_n|$ mit Hilfe von S(Y) konstruieren zu können. Enthält $\mathcal{Q}_n$ z.B. nur solche Permutationen π, für die gilt: $S(Y)=S(\pi Y)$, ist ein nichttrivialer Test offensichtlich unmöglich.

Als Beispiel sei noch einmal der zweifaktorielle vollständige Blockplan angeführt: Die Modellgleichung läßt sich abweichend von (3.69) auch schreiben als:

(3.88) $Y_{hij} = a_{ij} + b_h + \varepsilon_{hij}$ - h,i,j wie zuvor -

mit innerhalb der Blocke und als ganze Blöcke vertauschbaren Fehlern. Die Vertauschbarkeit der ganzen Blöcke kann auch noch aufgegeben werden.

Es soll die Wechselwirkungshypothese

$$H: \quad a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..} \qquad \text{für alle i,j}$$

getestet werden.

Die orthogonale Projektion von Y nach U^H ist (vgl. Roebruck, 1979):

$$(\hat{\bar{a}}_{i.} - \hat{\bar{a}}_{..}) + (\hat{\bar{a}}_{.j} - \hat{\bar{a}}_{..}) + \hat{\bar{a}}_{..} = \bar{Y}_{.i.} + \bar{Y}_{..j} - \bar{Y}_{...}$$

Damit ist die orthogonale Projektion $Y \to (U^H)^\perp$ komponentenweise:

(3.89) $$T_{hij} = Y_{hij} - \bar{Y}_{.i.} - \bar{Y}_{..j} + \bar{Y}_{...} \quad .$$

Diese ist eine Maximalinvariante.

Man sieht aber sofort, daß für diese Maximalinvariante nur ein trivialer Permutationstest resultieren kann. Da Mittelwerte über die Blöcke hinweg gebildet werden, können auch nur dieselben Zeilen oder Spalten über alle Blöcke hinweg vertauscht werden. Unabhängige Vertauschungen von Faktorstufen je Block und anschließende Transformation der Beobachtungen $T(\pi Y)$ können i.a. nicht durch eine Transformation, die Komponenten mit Mittelung über Blöcke hinweg enthält, und durch anschließende unabhängige Vertauschungen von Faktorstufen je Block $\pi(TY)$ 'kompensiert' werden. Außerdem sind Vertauschungen ganzer Blöcke zulässig. Diese drei Typen von Permutationen lassen aber die Teststatistik aus (3.71)

$$S(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} n(\bar{Y}_{.ij} - \bar{Y}_{.i.} - \bar{Y}_{..j} + \bar{Y}_{...})^2$$

invariant.

Der wesentliche 'Kniff' zur Konstruktion eines nichttrivialen Invarianz-Permutationstests liegt darin, sich mit einer lediglich invarianten Statistik

(3.90) $$T_{hij} = Y_{hij} - (\bar{Y}_{hi.} - \bar{Y}_{h..}) - (\bar{Y}_{h.j} - \bar{Y}_{h..}) - \bar{Y}_{h..}$$

wie in Abschnitt 3.1.5 (xii) zu 'begnügen'.

In diesem Fall ist zu erkennen, daß blockweise Vertauschungen der Faktorstufen je Faktor mit anschließender Transformation (3.90) durch Transformation und nachfolgende Permutation ' kompensierbar sind.

Die Teststatistik ist für beide Transformationen dieselbe. Bisher war von festen Blöcken ausgegangen worden. Zufallige Blockeffekte β_h sind aber ebenfalls kein Hindernis zur Konstruktion eines Invarianz-Permutationstests. Man muß fur die Verteilung der Blockeffekte nur annehmen, daß ihr Erwartungswert Null ist: $E(\beta_h)=0$. Die Modellgleichung wird dahingehend verändert, daß man nun eine Permutationsinvarianzannahme für die Verteilung F der Zufallsvariablen $\beta_h+\varepsilon_{hij}$ machen muß.

Eine sinnvolle Annahme ist die, daß gegenüber (3.88) die Vertauschbarkeit der Fehlerblöcke nicht mehr gegeben ist. An den Permutationsinvarianzeigenschaften von S(Y) ändert das nichts, da die in $\mathfrak{Q}_n$ noch zulässigen Vertauschungen ganzer Blöcke die Teststatistik invariant lassen.

Eine Konsequenz obiger Überlegungen ist, daß man auch für den zweifaktoriellen vollständigen Zufallsplan versuchen kann, invariante aber nicht maximalinvariante Transformationen zu suchen, die noch genugend Permutationen zur Bildung eines nichttrivialen Permutationstests übrig lassen .

Eine theoretisch zulässige Lösung wäre es, nach einer externen Variablen, von der kein Einfluß auf die abhängige Variable bekannt ist oder erwartet wird, die Beobachtungseinheiten zu schichten (stratifizieren), bevor sie den Faktorstufen(kombinationen) randomisiert zugeteilt werden. Dann fuhrt man die im CRF-rc noch zulässigen Vertauschungen der Faktorstufen je Faktor pro Stratum aus und bildet die Transformation der Y_{ijk} aus (3.22) lediglich je Stratum:

$$T_{ijkl}=Y_{ijkl}-\bar{Y}_{i..1}-\bar{Y}_{.j.1}+\bar{Y}_{...1} \qquad \begin{matrix} l=1,..,s \text{ Strata mit} \\ n_l \text{ Replikationen.} \end{matrix} \tag{3.91}$$

Als Teststatistik wählt man dann den zu (3.33) analogen Ausdruck

$$S(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} (\bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....})^2$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{T}^2_{ij..} \quad .$$

O.B.d.A. sei angenommen, daß in den Zellen des Planes je Stratum gleiche Zellbesetzungen vorliegen.

Da die Vertauschungen je Stratum unabhängig voneinander erfolgen, ist der auf S(Y) basierende Permutationstest nicht trivial. Welche Auswirkungen die Stratifizierung dagegen auf die Teststärke hat, ist nicht bekannt. Zu vermuten ist, daß sie in Abhängigkeit von der Heterogenität der Strata sinkt. Im Extremfall sind nur zwei Strata notwendig. Denkbar, jedoch mit i.a. negativer Auswirkung auf die Teststärke ist auch die Lösung, per Zufall Beobachtungseinheiten in $s \geq 2$ 'Pseudo-'Strata zusammenzufassen und dann wie zuvor beschrieben vorzugehen.

Dieselben Überlegungen sind direkt auf MANOVA-Probleme übertragbar, wie auch im Normalverteilungsfall zu allen exakten ANOVA-Tests selbst im gemischten Modell ein analoger MANOVA-Test existiert (vgl. Roebruck, 1982b). Im Abschnitt 4 wird die Analogie ausführlich dargestellt.

3.3 Randomisierungstests für lineare Hypothesen in linearen Modellen

3.3.1 Populationsmodell oder experimentelle Randomisierung

In Abschnitt 3.1 war ein sog. verteilungsfreies Populationsmodell betrachtet worden. Im Gegensatz zum üblichen linearen Modell mit Normalverteilungsannahme war dort nur als schwache Verteilungsannahme die Permutationsinvarianz der gemeinsamen Verteilung der Fehlerterme gegenüber bestimmten Untergruppen der symmetrischen Gruppe betrachtet worden. Insofern hängt die (statistische) Validität der Entscheidung über die jeweilige getestete lineare Hypothese davon ab, ob der Fehlerterm des tatsächlich erhobenen gesamten Beobachtungsvektors aus einer solchen Verteilung stammt. Falls es sich bei den Beobachtungen speziell um Realisationen unabhängiger, identisch verteilter Zufallsvariablen handelt, ist diese Bedingung stets erfüllt. Wie z.B. Wehrhahn & Ogawa (1982) oder John & Robinson (1983b) argumentieren, ist es bei der Annahme und Gültigkeit eines Modells mit unabhängigen, identisch verteilten Fehlervariablen überflüssig, eine experimentelle Randomisierung vorzunehmen. Da die Gültigkeit eines solchen Modells aber nicht überprüfbar ist, kommt der experimentellen Randomisierung, d.h. der zufälligen Zuordnung von Beobachtungseinheiten zu verschiedenen experimentellen Bedingungen bzw. der zufälligen Abfolge von verschiedenen experimentellen Bedingungen bei denselben Beobachtungseinheiten eine entscheidende Bedeutung zu (vgl. Bredenkamp, 1980, S. 18ff). Auf diesen Aspekt wird ausführlicher in Abschnitt 7 eingegangen.

Im Folgenden soll eine andere Betrachtungsweise ohne Verwendung eines Populationsmodells erläutert werden. Permutationstests lassen sich nämlich auch dann ausführen, wenn als einziges zufälliges Element, z.B. in einem linearen Modell, eine Zufallszuteilung der Beobachtungseinheiten zu den experimentellen Bedingungen angenommen wird. Diese Zuteilung wird i.a. durch diskrete sog. Design-Zufallsvariablen modelliert. Die zufällige Zuteilung muß in einem Experiment nicht bloß angenommen werden, sie ist (leicht) durch den Experimentator herstellbar. Permutationstests in einem Randomisierungsmodell werden auch als Randomisationstest oder Re-Randomisationstests (Gabriel & Hall, 1983) bezeichnet. Die Einflüsse der verschiedenen Stufen des (der) betrachteten Einflußfaktor(en) werden als additiv wirkende 'feste' Größen modelliert (Lehmann, 1959, S. 199). Mit jeder Beobachtungseinheit ist aber auch ein sog. 'unit error' (Einheiteneffekt bzw. Einheitenfehler, Wilk, 1955) verknüpft. Dieser repräsentiert die Unterschiedlichkeit mit der durch identische Faktorstufen gleich behandelte Beobachtungseinheiten 'reagieren' können. Die Hauptaufgabe der experimentellen Randomisierung ist es damit, die Einheitenfehler in einem statistischen Sinn zu kontrollieren. Rouanet, Bernard & Lecoutre (1986) versuchen eine mengentheoretische Darstellung ohne Zufallsvariablen.

Wenn, wie in der Psychologie üblich, Meßfehler in nicht unbeträchtlichem Umfang bei der Registrierung des (der) an den Beobachtungseinheiten zu erhebenden Meßwerte(s) auftreten, sind diese als ein zweites Zufallselement im Modell zu berücksichtigen. Diese auch als 'technical errors' (Wilk, 1955) bezeichneten Größen werden üblicherweise als Realisa-

tionen von um Null symmetrisch verteilte Zufallsvariablen aufgefaßt (vgl. John & Robinson, 1983b, Formel (3.1); Kempthorne, 1952, S. 123). Da der Vorgang der experimentellen Randomisierung i.a. unabhängig von dem Meßvorgang ist, können diese Fehler sinnvollerweise als unabhängig von den Design-Zufallsvariablen angenommen werden. In der Literatur (vgl. z.B. Ogawa, 1963, 1982) werden Modelle ohne Meßfehlerterm auch als 'Fisher-Modelle', diejenigen mit Meßfehler als 'Neyman-Modelle' bezeichnet. Eine andere Definition des Neyman-Modells ist bei Neyman (1935) und Holschuh (1978) zu finden.

Nicht in allen Fällen plausibel ist dagegen die Modellannahme, daß Behandlungseffekte und Effekte der Beobachtungseinheiten additiv sind. Es ist teilweise möglich, Tests auf Effekte von Faktoren auch dann sinnvoll auszuführen, wenn Einheiten x Faktoren-Interaktionen zugelassen werden (Rao, 1959, Zyskind, 1963; Scheffe, 1959, S. 318ff; Wilk, 1957; Wilk & Kempthorne, 1957). Dieser Fall soll im folgenden jedoch nicht betrachtet werden, wie auch für die Permutationstests in Abschnitt 3.1 implizit das Fehlen von solchen Wechselwirkungen durch die Art der Permutationsinvarianzstruktur in den Modellen angenommen worden war.

3.3.2 Verschiedene Darstellungen des Randomisierungsmodells

Alle wesentlichen Eigenschaften von Randomisierungsmodellen können bereits am vollständigen einfaktoriellen Blockplan veranschaulicht werden. Grundsätzlich soll Additivität zwischen experimentellen Einflußfaktoren und Beobachtungs-

einheiten angenommen werden (vgl. z.B., Zyskind, 1969, Kap. 4). Modelle mit Nicht-Additivität haben z.B. Zyskind (1963) oder Scheffe (1959, S. 318ff) analysiert. Wilk (1955) und Wilk & Kempthorne (1956, 1957) haben betont, daß es sich bei Randomisierungsmodellen um sog. abgeleitete (derived) finite lineare Modelle handelt. Die Charakterisierung als abgeleitetes Modell soll anzeigen, daß das verwendete statistische Modell vollständig durch die Art,in der das Experiment tatsächlich ausgeführt wurde, bestimmt wird (Addelman, 1970). Im Gegensatz dazu steht das sog. angenommene (assumed) infinite, übliche lineare Modell der Varianzanalyse (vgl. Kempthorne, 1977; Kempthorne & Doerfler, 1969, S. 234f).

In einem zufälligen vollständigen Blockplan werden k Faktorstufen anhand von kr Beobachtungseinheiten (experimentellen Einheiten) verglichen. Die Einheiten sind in r Blöcken mit je k Einheiten gruppiert. Je Block wird jede Einheit unter genau einer Faktorstufe beobachtet. Die Kombination von Strukturierung der Beobachtungseinheiten und des (der) experimentellen Faktors(en) wird in folgender 'strikt additiven' Modellgleichung (Kempthorne, 1955) für eine sog. konzeptuelle Beobachtung dargestellt (vgl. z.B. Zyskind, 1975; Kempthorne, 1952, Kap. 8.2):

$$y_{ij\nu} = u_{i\nu} + a_j \qquad \begin{array}{l} j=1,..,k \text{ Faktorstufen;} \\ i=1,..,r \text{ Blöcke; } \nu=1,..,k \end{array} .$$

(3.92) Die Strukturierung der Einheiten zeigt folgende Identität:

$$u_{i\nu} = \bar{u}_{..} + (\bar{u}_{i.}-\bar{u}_{..}) + (u_{i\nu}-\bar{u}_{i.}) \quad .$$

Eine tatsächliche Beobachtung in einem Block besteht nun aus einer zufälligen Zuordnung einer Einheit zu einer Faktorstufe unabhängig von den anderen Blöcken. Diese Zuordnung wird durch die diskreten Zufallsvariablen $\delta^j_{i\nu}$ die sog. Design-Variablen (Kempthorne, 1952; Wilk, 1955, S. 73) beschrieben. Für sie gilt:

$$(3.93)\qquad \delta^j_{i\nu} = \begin{cases} 1 & \text{in Block } i \text{ wird Faktorstufe } j \text{ auf Einheit } \nu \text{ angewendet} \\ 0 & \text{sonst} \end{cases}$$

mit

$$P(\delta^j_{i\nu}=1) = 1/k \quad \text{für alle } (i,j,\nu)\text{-Kombinationen};$$

$$P(\delta^j_{i\nu}\cdot\delta^{j'}_{i'\nu'}=1) = 1/k^2 \quad \text{falls } i\neq i'$$

und

$$P(\delta^{j'}_{i\nu'}=1 \mid \delta^j_{i\nu}=1) = \begin{cases} 1 & \nu=\nu', j=j' \\ 0 & \nu=\nu', j\neq j' \\ 0 & \nu\neq\nu', j=j' \\ 1/(k-1) & \nu\neq\nu', j\neq j' \end{cases}.$$

Die zufälligen Zuteilungen in den Blöcken erfolgen also unabhängig voneinander. Jede Faktorstufe tritt nur einmal pro Block auf, und jede Einheit wird pro Block nur einmal unter einer Faktorstufe beobachtet.

Für die tatsächliche Beobachtung y_{ij} gilt dann:

$$(3.94)\qquad y_{ij} = \sum_{\nu=1}^{k} \delta^j_{i\nu}\, y_{ij\nu} = \bar{u}_{..} + (\bar{u}_{i.} - \bar{u}_{..}) + \alpha_j + \sum_{\nu=1}^{k} \delta^j_{i\nu}\,(u_{i\nu} - \bar{u}_{i.}) .$$

Mit anderen Bezeichnungen hat man dann als abgeleitetes streng additives lineares Modell:

$$(3.95)\qquad y_{ij} = \mu + \beta_i + \alpha_j + e_{ij} \quad \text{mit } \beta_. = \alpha_. = 0 .$$

Die $\beta_i = \bar{u}_{i.} - \bar{u}_{..}$ sind als feste Blockeffekte interpretierbar, die e_{ij} als sog. Einheiten-Fehler oder Einheiten-Effekte (unit errors oder unit effects, vgl. Scheffe, 1959, S. 293; Kempthorne, 1977, S. 20). Dabei handelt es sich um den spezifischen Effekt der ν-ten Einheit in Block i unter der Faktorstufe j. Dieser 'Fehler' spiegelt die Eigenschaft der einzelnen Einheiten in einem Block wider, unter derselben

Faktorstufe potentiell abweichende Beobachtungen zu ergeben. Wichtig ist die folgende Beziehung für diese Einheiten-Fehler:

$$(3.96)\qquad \sum_{j=1}^{k} e_{ij} = \sum_{\nu=1}^{k} u_{i\nu} \sum_{j=1}^{k} \delta^{j}_{i\nu} - \sum_{j=1}^{k} \sum_{\nu=1}^{k} \delta^{j}_{i\nu}\, \bar{u}_{i.} \quad \text{für jeden Block } i \ .$$

Damit bestehen notwendigerweise Abhängigkeiten zwischen den Beobachtungen (vgl. z.B. White, 1975) und es gilt (vgl. Zyskind, 1969, S. 1363):

$$\mathrm{Var}(y_{ij}) = \mathrm{Var}(e_{ij}) = 1/k \sum_{\nu=1}^{k} (u_{i\nu}-\bar{u}_{i.})^2 =: \sigma_i \ , i=1,..,r$$

$$(3.97)\quad \text{und}\quad \mathrm{Cov}(y_{ij},y_{i',j'}) = \begin{cases} 0 & i\neq i' \\ -\sigma_i/(k-1) & j\neq j',\ i=i' \end{cases} \ .$$

Zwei verschiedene Beobachtungen im selben Block korrelieren also zu $\rho=-1/(k-1)$ (vgl. auch White, 1975). Falls man den gesamten $n=kr$-dimensionalen Beobachtungsvektor in der üblichen lexikographischen Schreibweise wie in Abschnitt 3.1.1 darstellt, läßt sich (3.97) unter der zusätzlichen Annahme $\sigma_i=\sigma$,$i=1,..,r$, darstellen als:

$$(3.98)\qquad \begin{aligned} \mathrm{Var}(y) = \mathrm{Var}(e) &= \sigma^2[(1-\rho)I_n + I_r \otimes J_k] \\ &= \sigma^2 \cdot k/(k-1) \cdot I_r \otimes (I_k - 1/k J_k) \ . \end{aligned}$$

Diese Varianz-Kovarianz Matrix ist notwendigerweise nicht von vollem Rang (vgl. Zyskind, 1967, S. 1092, 1100 oder auch Scariano & Davenport, 1984). Nimmt man zusätzlich (vgl. Kempthorne, 1953, Kap. 7.5) noch einen additiven unkorrelierten Meßfehler ε_{ij} in der Modellgleichung (3.95) an mit $E(\varepsilon_{ij})=0$ und $\mathrm{Var}(\varepsilon_{ij})=\sigma^2_\varepsilon$, hat man ein sog. Neyman-Modell mit der nichtsingulären Varianz-Kovarianz Matrix:

$$(3.99)\qquad \mathrm{Var}(y) = \sigma^2[(1-\rho)I_n + I_r \otimes J_k] + \sigma^2_\varepsilon I_n \ .$$

Pitman (1937b) differenziert nicht Einheiten- und Meßfehler.

Eine andere Darstellung des (Neyman-) Modells mit Meßfehler wählen Ogawa (1963, 1974, S. 85 ff) wie auch Wehrhahn & Ogawa (1982). In ihr werden die Beobachtungen in einer festen (lexikographischen) Reihenfolge notiert und die Zuordnungen von Einheiten und Blöcken und Faktorstufen zu Einheiten in Blöcken durch Indikatorvariablen angegeben:

$$(3.100)\quad y_f = \mu + \sum_{\nu=1}^{k} \delta_{\nu f}\cdot\alpha_\nu + \sum_{\nu'=1}^{k} \delta^*_{\nu' f}\cdot\beta_{\nu'} + u_f + \varepsilon_f$$

mit

$f=(i-1)k + j \quad ; i=1,..,r;\ j=1,..,k\ ;$

α_ν Effekt der Faktorstufe mit $\alpha_\cdot=0$,

$\beta_{\nu'}$ Effekt des Blockes ν' mit $\beta_\cdot=0$,

$$\delta_{\nu f}=\begin{cases}1 & \text{Beobachtungseinheit } f \text{ wird unter Faktorstufe } \nu \text{ beobachtet}\\ 0 & \text{sonst}\end{cases} \qquad \delta^*_{\nu' f}=\begin{cases}1 & \text{Beobachtungseinheit } f \text{ ist in Block } \nu'\\ 0 & \text{sonst}\end{cases}$$

Folgende Bedingungen gelten für die Einheitenfehler:

$\sum_{f=(i-1)k+1}^{ik} u_f = 0$ für jeden Block $i=1,..,r$,

und die Meßfehler sind unkorreliert und identisch (normal-) verteilt mit $E(\varepsilon_f)=0$ und $Var(\varepsilon_f)=\sigma^2_\varepsilon$.

In dem Modell (3.100) sind wie zuvor Wechselwirkungen zwischen Faktorstufen und Einheiten ausgeschlossen. Ansonsten hätte man statt u_f Ausdrücke der Form $\sum_{\nu=1}^{k} \delta_{\nu f} u_{\nu f}$ betrachten müssen.

Die Modellgleichung läßt sich mit Hilfe von aus den Indikatorvariablen gebildeten Inzidenzmatrizen $\delta_{n\times k}$ und $\delta^*_{n\times r}$ schreiben als:

$$(3.101)\quad y = \mu 1_n + \delta\alpha + \delta^*\beta + u + \varepsilon$$

mit

$\alpha = (\alpha_1,\ldots,\alpha_k)'$, $\beta = (\beta_1,\ldots,\beta_r)'$

und

$\delta^{*\prime} u = 0_r$, d.h. verschwindenden Einheitenfehlern je Block.

Die zufällige Zuteilung von Faktorstufen zu den Einheiten je Block läßt sich dann als (nxn)-dimensionale Permutationsmatrix $P=\bigoplus_{i=1}^{r} P_i \in \times_{i=1}^{r} S_k$ angewendet auf die Faktorstufeninzidenzmatrix δ als $p\delta$ darstellen. Unter Vernachlässigung von Meßfehlern hat y eine diskrete Wahrscheinlichkeitsverteilung, die bestimmt ist durch die $(k!)^r$ gleichwahrscheinlichen zulässigen Permutationen ausgedruckt durch P. Die zusätzliche Abhängigkeit von μ,β und den Einheitenfehlern u bleibt natürlich bestehen.

3.3.3 Tests linearer Hypothesen in Randomisierungsmodellen

Wie im Abschnitt zuvor, lassen sich alle wichtigen Überlegungen am Beispiel des einfaktoriellen vollständigen Blockplanes aufzeigen. Obwohl die e_{ij} in (3.95) bzw. die y_{ij} eine diskrete durch die Randomisierung induzierte Verteilung besitzen, wurde bereits von Pitman (1937a) und Welch (1937) die übliche F-Statistik, die im Normalverteilungsmodell gewisse Optimalitätseigenschaften besitzt, als Teststatistik auch im Randomisierungsmodell verwendet. Statt der F-Statistik selbst schlug Pitman die in monotoner Relation zu ihr stehende, im Normalverteilungsfall Beta-verteilte Teststatistik $U_{Faktor}=S_{Faktor}/(S_{Faktor}+S_{Fehler})$ vor. Obwohl er auch den in Abschnitt 2 beschriebenen exakten Test erwähnt und ihn für ein kleines Beispiel mit k=4 Faktorstufen und r=3 Blöcken ausführt, steht die Frage im Vordergrund, wie gut die diskrete Verteilung von U im Randomisierungsmodell mit der Betaverteilung im infiniten Normalverteilungsmodell übereinstimmt. Pitman bestimmt die ersten vier Momente der Verteilung von U unter dem Randomisierungsmodell und ver-

gleicht sie mit den entsprechenden Momenten der Beta-Verteilung mit (k-1)/2 und (k-1)(r-1)/2 Freiheitsgraden (sog. "principle of moments approximation", Kendall & Stuart, 1981). Es zeigt sich, daß nur die Erwartungswerte identisch sind (vgl. auch Scheffe, 1959, S. 318-325; Kempthorne et al, 1961, S. 174f) und die Varianzen in Abhängigkeit von der Variabilität der Blockvarianzen deutlich unterschiedlich ausfallen können. Der Unterschied verschwindet nahezu, falls die Blockvarianzen

$$s_i^2 := \sum_{j=1}^{k} y_{ij} - 1/k\left(\sum_{j=1}^{k} y_{ij}\right)^2$$

für alle i=1,..,r identisch ausfallen.

Unter diesen Bedingungen und bei zusätzlich 'nicht zu kleinem' r(k-1) gilt das auch für das 3. und 4. Moment. Falls die Blockvarianzen jedoch variieren, werden die Freiheitsgrade der Betaverteilung mit einem Korrekturfaktor

(3.102) $\phi = 1/(1-V/r) - 2/(r(k-1))$

mit

$$V = \left(r \sum_{i=1}^{r} s_i^4 / \left(\sum_{i=1}^{r} s_i^2\right)^2 - 1\right) \cdot (r/(r-1))$$

multipliziert, wenn der berechnete Teststatistikwert im Randomisierungsmodell zur Testentscheidung mit dem Schwellenwert im Normalverteilungsfall verglichen wird (vgl. Scheffe, 1959, S. 322; Ogawa, 1963, S. 1566; Ogawa, 1974, S. 104 und auch Lorenzen, 1984, S. 2609).

Gute Übereinstimmung in den ersten Momenten impliziert jedoch nicht gute Übereinstimmung in den oberen Quantilen der exakten und der Beta-Verteilung (vgl. Kempthorne et al., 1961, S. 175).

Einen weiteren wichtigen Unterschied betont auch bereits Scheffe (a.o.O., S. 325): Im Normalverteilungsfall sind die

verschiedenen mittleren Quadratsummen der ANOVA-Tabelle statistisch unabhängig und nur in diesem Fall ergibt sich ja auch bei Gültigkeit der jeweiligen Nullhypothese eine zentrale F-Verteilung für die jeweiligen Teststatistiken. Im Randomisierungstest sind MS_{Faktor} und MS_{Fehler} dagegen vollständig abhängig, denn

$$S_{Faktor} + S_{Fehler} = \sum_{ij}\sum (y_{ij}-\bar{y}_{i.})^2$$

ist unter den Vertauschungen innerhalb der Blöcke permutationsinvariant. Entsprechend erhält Wilk (1955, S. 75) auch eine Korrelation von - 1 zwischen diesen beiden Größen (vgl. auch Aastveit, 1983, S. 138 oder Okuno, 1962).
Deshalb liegt es nahe, wie zuvor bei den Invarianz-Permutationstests vorzugehen. Mit der Transformationsmatrix

$K=I_{rk}-1/kI_r \otimes J_k$ aus (3.10) erhält man wieder:

(3.103) $(Ky)_{ij} = y_{ij}-\bar{y}_{i.} = \alpha_j+e_{ij}$, da $\bar{e}_{i.}=1/n\sum_j e_{ij}=0$ für alle $i=1,..,r$.

unabhängig von den Blockeffekten. Bei Gültigkeit der Nullhypothese H: $\alpha_j=0$ sind die transformierten Beobachtungen frei von allen Modellparametern bis auf den Einheitenfehler. In der zuvor schon verwendeten Matrixschreibweise und mit $e=(e_{11},..,e_{1k},....,e_{r1},..,e_{rk})'$ und den Einheiteneffekten u aus (3.94) analog läßt sich (3.103) auch noch anders darstellen. Sei

(3.104) $\delta = ((\delta^j_{i\nu}))_{rk\times rk} = \bigoplus_{i=1}^{r} P_i$

die (Blockdiagonal-) Matrix der Design-Variablen mit den P_i als (kxk)-dimensionalen Permutationsmatrizen, d.h. jedes P_i hat in jeder seiner k Zeilen und Spalten genau eine 1 und sonst 0. Dann gilt für den gesamten Vektor der Einheiten-

fehler:

$$e = \left(\bigoplus_{i=1}^{r} P_i\right)\cdot(I_r \otimes (I_k - 1/kJ_k))\cdot u$$

(3.105)

$$= \left(\bigoplus_{i=1}^{r} P_i - 1/kJ_k\right)\cdot u$$

denn für (kxk)-dimensionale Permutationsmatrizen P gilt:

$$P\cdot I_k = P = I_k\cdot P \quad ; \quad P\cdot J_k = J_k = J_k\cdot P \quad .$$

Die Eliminierung der Blockeffekte erfolgt mit:

$$K = \bigoplus_{i=1}^{r} (I_k - 1/kJ_k) \quad .$$

Angewendet auf y und bei Gültigkeit der obigen Nullhypothese H ergibt das:

$$Ky = Ke = \left(\bigoplus_{i=1}^{r} (I_k - 1/kJ_k)\right)\cdot\left(\bigoplus_{i=1}^{r} (P_i - 1/kJ_k)\right)\cdot u$$

(3.106)

$$= \left(\bigoplus_{i=1}^{r} (P_i - 1/kJ_k)\right)\cdot u = e \quad .$$

Bei Gültigkeit der Nullhypothese haben im Randomisierungsmodell die transformierten Beobachtungen dieselbe diskrete Verteilung wie die Einheitenfehler. In Abhängigkeit von dem im jeweiligen Experiment tatsächlich vorhandenen Einheiteneffektvektor u hat jede Realisation von δ dieselbe Wahrscheinlichkeit von $1/(k!)^r$. Die Verteilung von $\delta\cdot u$ ist $\times_{i=1}^{r} S_k$-invariant. Beim Invarianzpermutationstest in Abschnitt 3.1.4 galt unter H dagegen 'lediglich' $Ky=K\varepsilon$. Dort mußte überlegt werden, für welche Permutationen $\pi\in S_{rk}$ zusätzlich gilt $\pi K=K\pi$.

Den Unterschied zwischen Randomisierungsmodell und Invarianz-Permutationstest Modell kann man sich auch noch so verdeutlichen:

Im Randomisierungsmodell gilt, wie (3.97) schon zeigte:

$$Cov(e_{ij}, e_{i'j'}) = 0 \quad \text{falls } i\neq i' \text{ ist} \quad .$$

Mit den Methoden aus (3.1.4 (xi)) rechnet man für den Permutationstest aus, wobei $KY=:T$ gesetzt sei (unter H ist $t_{ij}=e_{ij}$!) :

$$P_{rk}(T_{ij}=x) = 1/rk$$

$$P_{rk}(T_{ij}=x;T_{i'j'}=x') = \begin{cases} 1/rk(k-1) & i=i',\ j\neq j' \\ 1/r(r-1)k & i\neq i',\ j=j' \\ 1/(r(r-1)k(k-1)) & i\neq i',\ j\neq j' \end{cases}$$

$$E_{P_{rk}}(T_{ij}) = 1/rk\sum_{ij}\sum t_{ij} = 0$$

$$Var_{P_{rk}}(T_{ij}) = 1/rk\sum_{ij}\sum t_{ij}^2 =: ((r-1)(k-1)/rk)\sigma^2_{rk,t}$$

(3.107)
$$Cov_{P_{rk}}(T_{ij},T_{ij'}) = (-(r-1)/rk)\sigma^2_{rk,t}$$

$$Cov_{P_{rk}}(T_{ij},T_{i'j}) = (-(k-1)/rk)\sigma^2_{rk,t}$$

$$Cov_{P_{rk}}(T_{ij},T_{i'j'}) = (1/rk)\sigma^2_{rk,t}$$

und damit

$$Var(T) = \sigma^2_{rk,t}(I_r-1/rJ_r)\otimes(I_k-1/kJ_k)$$

(vgl. auch Nelder (1954, 1964a)).

Es bestehen unter der Modellannahme $\varepsilon\sim F(x)$ mit $F(\pi x)=F(x)$ für alle $\pi\in \underset{i=1}{\overset{r}{\times}} S_k$ oder auch $\pi\in S_k\sim S_r$ unter der Nullhypothese von Null verschiedene Kovarianzen für Werte aus unterschiedlichen Blöcken im Gegensatz zu (3.98). Für wachsende Blockanzahl r streben die Korrelationen für verschiedene Blöcke jedoch gegen Null.

Im einfaktoriellen Blockplan gelangt man aber in beiden Modellen (Randomisierungsmodell bzw. Populationsmodell mit schwachen Verteilungsvoraussetzungen) für einen exakten Permutationstest zu derselben Teststatistik, die dem Zähler der ANOVA-Teststatistik analog ist. Im Randomisierungsmodell entspricht der Fall gleicher Blockvarianzen $\sigma_i=\sigma$ dem der Permutationsgruppe $Q^*_{rk}=S_k\sim S_r$ für den Fehlervektor im Populationsmodell. Wie Zyskind (1969, S. 1364 unten, 1975) aber für das Randomisierungsmodell betont, ist auch für ungleiche σ_i und entsprechend für $Q^{**}_{rk}=\underset{i=1}{\overset{r}{\times}} S_k$ ein (unverzerrter) Test möglich. Das gilt, obwohl die linearen Schätzer der Effekte

im Modell nicht mehr blue-Schätzer (best linear unbiased estimate) sind (vgl. Zyskind, 1967, 1969, S. 1364). Den Fall mit Block-Faktor Interaktionen analysiert Scheffe (1959, S. 292ff) und Robinson (1972) den Fall unvollständiger Blockpläne.

Wie am Beispiel des einfaktoriellen Blockplanes demonstriert, gilt bei invarianter Reduktion des Testproblems im Randomisierungsmodell Ky=e und nicht lediglich Ky=Ke bei fehlendem Einfluß des experimentellen Faktors. Es ist also eigentlich wieder für jeden der bisher behandelten Versuchspläne zu prüfen, wie das Transformationsverhalten der Fehler nach invarianter Transformation zur Eliminierung störender Effekte im Randomisierungsmodell unter der jeweiligen Nullhypothese aussieht.

3.3.4 Existenz von Invarianz-Permutationstests in Randomisierungsmodellen

In der Literatur ist das Problem der Herleitung von geeigneten Teststatistiken für Randomisierungstests in komplexeren, mehrfaktoriellen Versuchsplänen nicht sehr intensiv behandelt worden. Im Vordergrund standen vielmehr folgende Gesichtspunkte (vgl. besonders Wilk & Kempthorne, 1955; Zyskind, 1962, 1963, 1969, 1975; Lorenzen, 1984):

(i) Angabe eines finiten Populationsmodells wie in (3.92) für den vollständigen einfaktoriellen Blockplan bestehend aus additiv verknüpften Einflußgrößen für die Faktorstufen(kombinationen) sowie Effekte der Beobachtungseinheiten und eventuell einen Term für Meßfehler.

(ii) Bestimmung eines sog. abgeleiteten linearen Modells wie in (3.94), das die beobachteten Werte darstellt über durch sog. elementare Design-Variablen δ verknüpfte konzeptuelle Beobachtungen aus (i). Die probabilistischen Eigenschaften des Modells sind im wesentlichen bestimmt durch die Verteilungseigenschaften der diskreten Design-Variablen.

(iii) Analyse der Erwartungswerte unter dem Randomisierungsmodell von zur ANOVA mit Normalverteilungs-Annahme analog gebildeten mittleren Quadratsummen insbesondere unter dem Aspekt der unverzerrten Schätzung von Effekten. Es wurden auch sehr allgemeine Bedingungen dafür angegeben, wann die üblichen Kleinste-Quadrat-Schätzer der linearen Modellparameter gleichzeitig beste lineare unverfälschte Schätzer in den Randomisierungsmodellen mit singulärer Varianz-Kovarianz Matrix des gesamten Beobachtungsvektors darstellen (Zyskind, 1969).

Als Teststatistik wird häufig die übliche F-Statistik oder die in monotoner Beziehung dazu stehende Statistik $U = SS_{Effekt} / (SS_{Effekt} + SS_{Fehler})$ verwendet. Diese hat den Vorteil, daß der Nenner unter dem jeweiligen Randomisierungsmodell eine Konstante ist. Scheffe (1959), Wilk (1955) und Okuno (1958, 1962) betonen, daß im Unterschied zum üblichen ANOVA- Modell die verschiedenen Quadratsummen (als Zufallsvariablen) notwendigerweise aufgrund des Randomisierungsmodells miteinander (negativ) korreliert sind.

Die Begründungen für die Verwendung von F- oder U-Statistiken als Teststatistiken sind folgende:

(i) Der Vergleich der ersten 2 - 4 Momente der Verteilung von F oder U unter dem Randomisierungsmodell bzw. dem ANOVA-Modell zeigt Übereinstimmung im Erwartungswert und ähnliche Werte für die anderen Momente, allerdings in Abhängigkeit von den Einheiten-Effekten. Es lassen sich auch Korrekturfaktoren für die Freiheitsgrade der parametrischen Statistiken angeben, die diese Ähnlichkeit vergrößern.

(ii) Asymptotische Berechnungen für wachsenden Stichprobenumfang zeigen, daß bei recht schwachen Bedingungen an die Beobachtungen (vgl. z.B. Wald & Wolfowitz, 1944; Silvey, 1954) im Normalverteilungs- und im Randomisierungsfall dieselbe asymptotische Verteilung der Teststatistik resultiert.

(iii) Es gibt verschiedene numerische Beispiele (vgl. z.B. Scheffe, 1959, Fußnote S. 324), in denen die exakte oder approximativ bestimmte Permutationsverteilung der Teststatistik bzw. deren oberes $(1-\alpha)$ -Quantil 'gut' mit dem entsprechenden Quantil einer F-oder Beta-Verteilung übereinstimmt. Es gibt aber auch Fälle, in denen ein deutlicher Unterschied bei der Reanalyse von Daten in einem 2-Zeitpunkte Cross-Over Plan resultiert (Gomez-Marin & McHugh, 1984).

Eine theoretisch zufriedenstellende Argumentation ist mit dieser Begründung kaum möglich. Überträgt man den Ansatz über Invarianz-Permutationstests aus den Abschnitten 3.1 und 3.2 auf das Randomisierungsmodell, zeigt das Beispiel des einfaktoriellen Blockplanes aus dem letzten Abschnitt nicht zwingend, daß man zu identischen Ergebnissen wie im Populationsmodell mit 'schwachen' Vertauschbarkeitsannahmen an den

Fehlervektor gelangen wird. Die Existenz bzw. Nichtexistenz eines Invarianz-Permutationstests im Randomisierungsmodell auf Wechselwirkung soll abschließend für den Split-Plot Plan bzw. den zwei-faktoriellen vollständigen Zufallsplan veranschaulicht werden.

Im Folgenden werden zusätzliche additive, voneinander und von den Einheitenfehlern unabhängige Meßfehler mit Erwartungswert Null, obwohl in der psychologischen Forschung sicher von großer Bedeutung, nicht in den Modellen berücksichtigt, da bei obigen Voraussetzungen der gesamte n-dimensionale Meßfehlervektor stets eine S_n-invariante Verteilung besitzt und die Permutations-Invarianzeigenschaften der (transformierten) Einheitenfehler stets höchstens eingeschränkter sein können.

Im abgeleiteten linearen Modell für den Split-Plot Plan gibt es zwei verschiedene durch die Art der experimentellen Randomisierung induzierte Fehlerterme. Das Modell läßt sich ohne Berücksichtigung von Meßfehlern darstellen als (vgl. Kempthorne, 1952; Kempthorne et al., 1961, S. 160ff; Lorenzen, 1984; Zyskind, 1969, 1975):

Als konzeptuelle Beobachtung hat man bei Annahme gleichvieler Replikationen je Plot-Faktorstufe:

$$y_{ijk\nu v} = u_{i\nu v} + \alpha_j + \beta_k + (\alpha\beta)_{jk} \qquad \begin{matrix} j=1,..,r \text{ Plotfaktorstufen} \\ k=1,..,c \text{ Splitfaktorstufen} \\ i=1,..,n \text{ Blöcke je Plotstufe} \end{matrix}$$

(3.108) mit folgender Strukturierung der Einheitenwerte:

$$u_{i\nu v} = \bar{u}_{...} + (\bar{u}_{i..} - \bar{u}_{...}) + (\bar{u}_{i\nu .} - \bar{u}_{i..}) + (u_{i\nu v} - \bar{u}_{i\nu .})$$

$$\nu=1,..,r;\ v=1,..,c \quad .$$

Um Vergleiche mit den Ergebnissen aus der Literatur zu erleichtern, wurde die Abfolge der Indizes wie bei Zyskind (1969, 1975) gewählt. Eine tatsächliche Beobachtung in einem Block (d.h. eine Replikation unter der Plotfaktorstufe j) besteht nun aus einer zufälligen Zuordnung einer Plot-Faktorstufe zu einem gesamten Block, und innerhalb dieses Blocks wird eine Einheit (i,ν,v) zufällig zu einer Split-Faktorstufe zugewiesen unabhängig von den anderen Blocken. Diese beiden Zuordnungen werden durch die beiden unabhängigen diskreten Design-Zufallsvariablen $\delta^{j}_{i\nu}$ und $\delta^{*k}_{i\nu v}$ beschrieben mit:

(3.109)

$$\delta^{j}_{i\nu} = \begin{cases} 1 & \text{In Block } i \text{ wird Plot-Faktorstufe } j \text{ auf die } c \text{ Einheiten von 'whole Plot' } \nu \text{ angewendet} \\ 0 & \text{sonst} \end{cases}$$

$$\delta^{*k}_{i\nu v} = \begin{cases} 1 & \text{in Block } i \text{ wird Split-Faktorstufe } k \text{ auf die Einheit } v \text{ in whole Plot } \nu \text{ angewendet} \\ 0 & \text{sonst} \end{cases}$$

Für eine tatsächliche Beobachtung y_{ijk} gilt dann folgendes abgeleitete lineare Modell:

(3.110)

$$\begin{aligned} y_{ijk} &= \sum_{\nu=1}^{r} \sum_{v=1}^{c} \delta^{j}_{i\nu} \delta^{*k}_{i\nu v} y_{ijk\nu v} \\ &= \bar{u}_{...} + (\bar{u}_{i..} - \bar{u}_{...}) + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \\ &\quad + \sum_{\nu=1}^{r} \delta^{j}_{i\nu} (\bar{u}_{i\nu.} - \bar{u}_{i..}) + \sum_{\nu=1}^{r} \sum_{v=1}^{c} \delta^{j}_{i\nu} \delta^{*k}_{i\nu v} (u_{i\nu v} - \bar{u}_{i\nu.}) \; . \end{aligned}$$

Mit anderen Bezeichnungen ergibt das:

(3.111) $y_{ijk} = \mu+\pi_i+\alpha_j+\beta_k+(\alpha\beta)_{jk}+e_{ij}+e^*_{ijk}$

mit $\alpha_.=\beta_.=(\alpha\beta)_{i.}=(\alpha\beta)_{.j}=\pi_.=0$.

Die $\pi_i:=\bar{u}_{i..}-\bar{u}_{...}$ sind als Whole-plot - bzw. Blockeffekte, die e_{ij} als Blockfehler und die e^*_{ijk} als Einheitenfehler je Block interpretierbar.

Für die beiden Fehlerterme gelten die Abhängigkeiten erzeugenden Bedingungen:

(3.112) $\sum_{j=1}^{r} e_{ij} = 0$ für jeden Block i

$\sum_{k=1}^{c} e^*_{ijk} = 0$ für jede Block-Plotfaktor Kombination (i,j)

In Matrixschreibweise hat man für die Design-Variablen:

(3.113) $\delta = ((\delta^j_{i\nu}))_{nrc\times nrc} = \bigotimes_{i=1}^{n} (P_i\otimes I_c)$ mit P_i (r×r)-dimensionaler Permutationsmatrix

$\delta^* = ((\delta^{*k}_{i\nu\nu}))_{nrc\times nrc} = \bigotimes_{i=1}^{n}\bigotimes_{j=1}^{k} P^*_{ij}$ mit P^*_{ij}(c×c)-dimensionaler Permutationsmatrix.

Die beiden gesamten 'Einheiten'-Fehlerterme des abgeleiteten linearen Modells lassen sich bei lexikographischer Anordnung der Komponenten schreiben als:

$$
\begin{aligned}
e &= (\bigotimes_{i=1}^{n}(P_i\otimes I_c))(I_{nr}\otimes 1/cJ_c - I_n\otimes 1/rcJ_{rc})\cdot u \\
&= (\bigotimes_{i=1}^{n}((P_i-1/rJ_r)\otimes(1/cJ_c)\cdot u \\
e^* &= ((\bigotimes_{i=1}^{n}(P_i\otimes I_c))\cdot(\bigotimes_{i=1}^{n}\bigotimes_{j=1}^{r}P^*_{ij})(\bigotimes_{i=1}^{n}\bigotimes_{j=1}^{r}(I_c-1/cJ_c)))\cdot u \\
&= (\bigotimes_{i=1}^{n}(P_i\otimes(P^*_{ij}-1/cJ_c)))\cdot u \quad .
\end{aligned}
\tag{3.114}
$$

Benutzt man die die störenden Parameter eliminierende Transformationsmatrix $K=(I_{nr}-1/nrJ_{nr})\otimes(I_c-1/cJ_c)$ aus (3.64), erhält man mit n.=nr bei gleichen Stichprobenumfängen unter

der Nullhypothese H: $(\alpha\beta)_{jk}=0$ für alle j, k:

(3.115) $Ky = Ke^*$

unabhängig von den Blockfehlern, denn Ke = 0, was man durch Ausrechnen leicht überprüft. Im Gegensatz zum einfaktoriellen Blockplan gilt nicht $Ke^*=e^*$ sondern:

$$Ke^* = e^* - (1/nrJ_{nr}\otimes I_c)\cdot e^* \text{ , denn}$$

(3.116) $$-(I_{nr}-1/nrJ_{nr})\otimes 1/cJ_c\cdot e^* = 0$$

wegen der zweiten Zeile in (3.112).

K und $(I_{nrc}-1/nrJ_{nr}\otimes I_c) = (I_{nr}-1/nrJ_{nr})\otimes I_c$ haben aber dieselben $Q_{nr}=S_{nr}\otimes S_c$ - Permutationseigenschaften wie in Abschnitt 3.1.4 (xi).

Da $\delta^* S_{nr}\sim S_c$ invariant ist wie auch die Fehlerverteilung im Populationsmodell mit schwachen Verteilungsannahmen, ist die Teststatistik aus (3.65) unter der Nullhypothese in beiden Fällen unter allen $(nr)!/(n!)^r$ Vertauschungen ganzer Blöcke für den Test auf Wechselwirkungen zu betrachten.

Im zweifaktoriellen vollständigen Zufallsplan (mit gleicher Zellbesetzung) hat man im abgeleiteten Modell als Beobachtungen (vgl. Wilk & Kempthorne, 1955).

(3.117) $$y_{ijk} = \sum_{\nu=1}^{N} \delta_\nu^{ijk}\, Y_{ijk\nu} \quad i=1,..,r;j=1,..,c;k=1,..,n;N=rcn$$

mit $$\delta_\nu^{ijk} = \begin{cases} 1 & \text{k-te Replikation der Faktorstufenkombination (i,j) wird auf Einheit } \nu \text{ angewendet} \\ 0 & \text{sonst} \end{cases}$$

mit $Y_{ijk} = \alpha_i+\beta_j+(\alpha\beta)_{ij}+u_\nu$ und $\alpha_\cdot=\beta_\cdot=(\alpha\beta)_{i\cdot}=(\alpha\beta)_{\cdot j}= 0$ und der 'Strukturierung' der Einheiteneffekte:

$$u = \bar{u}_\cdot + (u_\nu - \bar{u}_\cdot) \text{ , sodaß}$$

$$y_{ijk} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\underbrace{\sum_{\nu=1}^{N} \delta_\nu^{ijk}(u_\nu-\bar{u}_\cdot)}_{=: e_{ijk}} \quad \text{mit} \sum\sum\sum_{ijk} e_{ijk} = 0 .$$

Im (zweifaktoriellen) vollständigen Zufallsplan fehlt damit eine echte Strukturierung der Einheiten.

In Matrixschreibweise hat man mit der S_N-invarianten Design-Variablenmatrix $\delta=((\delta_\nu^{ijk}))_{N\times N}=:P$

$$e = (P-1/NJ_N)\cdot u \quad .$$

Obige Nebenbedingung an die Einheitenfehler ist dann $1/NJ_N\cdot e= 0$.

Mit der Transformationsmatrix K aus (3.31) hat man bei Gültigkeit der Nullhypothese H : $(\alpha\beta)_{ij}=0$ für alle i,j:

(3.118) $Ky = Ke = e-((I_r\otimes 1/cJ_c+1/rJ_r\otimes I_c)\otimes 1/nJ_n)\cdot e$.

K und die Matrix rechts in der obigen Gleichung sind aber permutationsinvariant gegenüber $Q_N=S_n\wr(S_r\otimes S_c)$ aus (3.32). Man ist also in derselben Situation wie zuvor, daß kein nichttrivialer Invarianz-Permutationstest auch für das Randomisierungsmodell existiert.

Auch für die übrigen in Abschnitt 3.1.4 behandelten Tests auf Wechselwirkung in mehrfaktoriellen Versuchsplänen gilt, daß für den(die) entsprechenden Einheitenfehlervektor (en) in keinem Fall Ke = e ist. Die Transformationsmatrix K zur invarianten Reduktion des Testproblems stiftet zusätzliche Abhängigkeiten zwischen den Einheitenfehlern. Zwar verschwinden wie in (3.116) oder (3.118) einzelne Teile von K aufgrund der linearen Nebenbedingungen an die Einheitenfehler. Die nichtverschwindenden Anteile haben aber stets dieselbe Permutationsinvarianzstruktur wie K selbst.

Unter Verwendung eines etwas allgemeineren Modells als in dieser Arbeit demonstriert auch Lorenzen (1984) die Problematik der Verwendung eines F-Tests, wenn Einheiten-Effekte vorliegen. Für ein- und mehrfaktorielle Pläne kann

er zeigen, daß bei vorhandenen Einheitenfehlern die üblichen ANOVA-Teststatistiken auch bei Gültigkeit der jeweiligen Nullhypothese nichtzentrale F-Verteilungen aufweisen.

Wilk & Kempthorne (1955, 1956) untersuchen abgeleitete lineare Modelle für ein- und mehrfaktorielle vollständige Zufallspläne, in denen Additivität zwischen experimentellen Faktoren und Einheiten nicht notwendigerweise gelten muß.

Die Diskussion über den Stellenwert von Randomisierungsmodellen wird in Abschnitt 7 wieder aufgenommen.

4. INVARIANZ-PERMUTATIONSTESTS FÜR LINEARE HYPOTHESEN IN LINEAREN MODELLEN - MULTIVARIATER FALL

4.1 Multivariate Experimente

In einem multivariaten Experiment wird der Einfluß einer oder mehrerer unabhängiger Variablen (Faktoren) auf mehrere abhängige Variablen - erhoben an den Versuchseinheiten - überprüft. Eine multivariate Analysemethode sollte vornehmlich dann verwendet werden, wenn die Einflüsse der unabhängigen Variable(n) auf eine Menge von abhängigen Variablen von Interesse sind (vgl. Bredenkamp, 1980; Hager & Westermann, 1983a; Bray & Maxwell, 1982; Meredith et al., 1974). Das ist häufig der Fall, wenn mehrere verschiedene (Gruppen von) abhängige(n) Variable(n) als Indikatoren für ein psychologisches Konstrukt angesehen werden. Die experimentelle(n) Hypothese(n) sollten sich ebenfalls auf Einflüsse des(der) Faktors(en) auf alle erhobenen (Gruppen von) abhängigen Variablen beziehen. Auch unter dem Gesichtspunkt einer möglichst strengen Prüfung einer Kausalhypothese ist es sinnvoll, mehrere - möglichst unterschiedliche - Operationalisierungen der theoretischen Begriffe (Konstrukte) zu berücksichtigen (vgl. Hager & Westermann, 1983a, S. 46).

Gegenüber mehreren univariaten Analysen werden die Zusammenhänge zwischen den abhängigen Variablen mitberücksichtigt. Es ist u.U. möglich, daß erst bei multivariater Betrachtung signifikante Effekte feststellbar sind (vgl. z.B. Ahrens & Läuter, 1981; Edgington, 1980, S. 181).

Die Bedeutung multivariater Verfahren für die (experimentelle) Psychologie ist relativ früh erkannt worden; siehe z.B. das Handbuch von Cattell (1966). Bock (1966) bezeichnet z.B. die pädagogische Psychologie als 'a field where research problems are inherently multivariate' (S. 822). Bortz (1977, S. 579ff, S. 697) gibt mehrere Beispiele für psychologische Konstrukte, die nur anhand mehrerer Indikatoren sinnvoll erfaßt werden können.
In der Regel ist die Überprüfung eines "globalen" Haupt- oder Wechselwirkungseffektes für alle abhängigen Variablen zusammengenommen nur der erste Schritt. Zur detaillierten Analyse dieser Effekte gibt es eine Vielzahl von verschiedenen Strategien (vgl. z.B. Bray & Maxwell, 1982; Gabriel, 1968, 1969b). Auf diese wird in den Abschnitten 5.3 und 5.5. eingegangen.

4.2. Ansätze zu Permutationstests für multivariate Experimente

Eine Übertragung des Permutationstestprinzips auf den multivariaten Fall ist konzeptuell einfach (vgl. Puri & Sen, 1971, S. 68; Bell & Smith, 1969). Anstelle der n univariaten Beobachtungen aus Kapitel 2 treten nun n p>1-dimensionale Beobachtungsvektoren $x_i=(x_{i1},\ldots,x_{ip})$.
Unter der jeweiligen Nullhypothese ist die n·p-dimensionale Verteilungsfunktion F(x) der n·p-dimensionalen Zufallsvektoren X_i - u.U. nach einer linearen Transformation - invariant unter einer Gruppe $\tilde{\mathcal{G}}_n$ von Permutationen der p-dimensionalen Vektoren, sog. Vektorpermutationen (vgl. Edgington, 1980, S. 190ff); anders

formuliert: Unter H_o haben alle zulässigen Vektor-Permutationen dieselbe Wahrscheinlichkeit. Für $\hat{\mathcal{Q}}_n$ gilt: $\hat{\mathcal{Q}}_n = \mathcal{Q}_n \otimes \{i_p\}$, d.h. $\hat{\mathcal{Q}}_n$ is isomorph zu der entsprechenden Permutationsgruppe $\mathcal{Q}_n$ von S_n von p=1-dimensionalen Zufallsvariablen. Durch die Vertauschung der jeweils ganzen Zufallsvektoren X_i bleibt die Abhängigkeitsstruktur unter den abhängigen Variablen erhalten und F ist unter H_o symmetrisch bezüglich seiner p-dimensionalen Vektorkomponenten. Theoretisch interessant ist, daß auch multivariate Permutationstests gemäß Konstruktion stets verteilungsfrei sind. Für viele multivariate Verallgemeinerungen von univariaten Rangtests gilt diese Eigenschaft nicht. Selbst bei Gültigkeit der betreffenden Nullhypothese hängen die Tests von der zugrundeliegenden unbekannten Verteilung ab (Puri & Sen, 1971, S. 69, 183). Auch die Formulierung von (Null-)Hypothesen z.B. in linearen Modellen erfolgt i.a. durch Ersetzung der univariaten Parameter durch p-dimensionale Vektor-Parameter.

Gelöst werden muß das Problem der Wahl einer geeigneten Teststatistik für das betreffende Testproblem.

(i) Wald & Wolfowitz (1944) wählen für das unverbundene Zweistichprobenproblem (mit p=2) eine Teststatistik in Analogie zum parametrischen T^2-Test von Hotelling (1931). Sei das Experiment beschrieben durch

$$Y = (Y'_{11},..,Y'_{1n_1};Y'_{21},...,Y'_{2n_2})'$$

mit $Y_{ij}=(Y_{ij1},Y_{ij2})'$ $i=1,2$; $j=1,..,n_i$; $n_1+n_2=n$

und Y_{ij} verteilt nach $F_i(x)$ mit $F_1(x_1,x_2)=F_2(x_1-\delta_1,x_2-\delta_2)$

und $\delta=(\delta_1,\delta_2)'$.

Die (lineare) Nullhypothese lautet in Analogie zu Abschnitt 3.1.5 (i):

$H: \delta=0$ gegen $A: \delta\neq 0$.

Unter H ist die Verteilungsfunktion von Y $F(x), x\in R^{2n}$, invariant unter allen Vektor-Permutationen $\pi\in\hat{S}_n=S_n\otimes\{i_2\}$.

Eine mögliche Teststatistik, die eine direkte Verallgemeinerung der univariaten Teststatistik darstellt, ist die quadratische Form für $p>1$ beliebig, vgl. z.B. Bortz, 1977, S. 702f):

$$(4.1)\quad T^2 = n_1n_2/(n_1+n_2)\cdot(\bar{Y}_{1.}-\bar{Y}_{2.})'V_n^{-1}(\bar{Y}_{1.}-\bar{Y}_{2.})$$

$$= n_1n_2/(n_1+n_2)\sum_{l=1}^{p}\sum_{l'=1}^{p}(\bar{Y}_{1.1}-\bar{Y}_{2.1})\sigma^{*}_{ll'}(\bar{Y}_{1.1'}-\bar{Y}_{2.1'})$$

mit

$$V_n:=((\sigma_{ll'}))=((1/(n_1+n_2-2)\{\sum_{j=1}^{n_1}(Y_{1jl}-\bar{Y}_{1.l})(Y_{1jl'}-\bar{Y}_{1.l'})+$$

und $((\sigma^{*}_{ll'}))_{p\times p}=V_n^{-1}$ $\qquad + \sum_{j=1}^{n_2}(Y_{2jl}-\bar{Y}_{2.l})(Y_{2jl'}-\bar{Y}_{2.l'})\}))_{p\times p}$

Eine für die Durchführung eines Permutationstests geeignetere Teststatistik ist in Analogie zu (3.37):

$$(4.2)\quad \hat{T}^2 = (\bar{Y}_{1.}-\bar{Y}_{2.})'\hat{V}_n^{-1}(\bar{Y}_{1.}-\bar{Y}_{2.}) = \sum_{l=1}^{p}\sum_{l'=1}^{p}\tilde{\sigma}^{*}_{ll'}(\bar{Y}_{1.1}-\bar{Y}_{2.1})(\bar{Y}_{1.1'}-\bar{Y}_{2.1'})$$

mit

$$\hat{V}_n = ((\tilde{\sigma}_{ll'}))_{p\times p}$$

$$= (((n_1+n_2)/(n_1n_2(n_1+n_2-2)))\cdot\{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ijl}-\bar{Y}_{..l})(Y_{ijl'}-\bar{Y}_{..l'})\}))$$

und

$$((\tilde{\sigma}^{*}_{ll'}))_{p\times p} = \hat{V}_n^{-1} \quad .$$

(4.2) hat gegenüber (4.1) den im Folgenden oft genutzten Vorteil, daß $((\tilde{\sigma}_{ll'}))$ im Gegensatz zu $((\sigma_{ll'}))$ invariant ist für die $(n_1+n_2)!$ unter der Nullhypothese zulässigen Permutationen der p-dimensionalen Beobachtungsvektoren. Damit muß zur Durchführung des exakten Permutationstests die (verallgemeinerte) Inverse der Varianz-Kovarianz Matrix nur einmal gebildet werden. Weiterhin können wieder nur die $(n_1+n_2)!/n_1!n_2!$ Permutationen von Beobachtungsvektoren über die beiden Stichproben hinweg zu unterschiedlichen Prüf-

größenwerten von (4.2) führen. Die Wahl der modifizierten Teststatistik ist jedoch nur sinnvoll, weil man zeigen kann, daß T^2 und $\tilde{T}^2$ in monotoner Beziehung zueinander stehen und damit der Permutationstest für beide stets dieselben p-Werte ergibt. Wald & Wolfowitz behandeln nur den asymptotischen Fall und zeigen, daß $\tilde{T}^2$ für $n\to\infty$ gegen eine χ^2-Verteilung mit df=min $(2,rg(\tilde{V}_n))$ konvergiert.
Maritz (1981) analysiert einen bivariaten Mediantest.

(ii) Für das p-dimensionale verbundene Zweistichprobenproblem gibt Arnold (1964) einen Permutationstest an. Seien in Verallgemeinerung von 3.1.5 (ii) die $Y_i=Y_{1i}-Y_{2i}, i=1,..,n$, p-dimensionale Differenzenvektoren verteilt nach p-dimensionalem symmetrischem $F_i(x-\delta)$
Die Nullhypothese lautet:

H: $\delta=0$ gegen A: $\delta\neq 0$.

Als Teststatistik hat man die quadratische Form:

(4.3) $$T^2 = n(\bar{Y}_{1.}-\bar{Y}_{2.})'V_n^{-1}(\bar{Y}_{1.}-\bar{Y}_{2.})$$

$$= n\sum_{l=1}^{p}\sum_{l'=1}^{p}(\bar{Y}_{1\cdot l}-\bar{Y}_{2\cdot l})\,\sigma^{*}_{ll'}\,(\bar{Y}_{1\cdot l'}-\bar{Y}_{2\cdot l'})$$

mit

$$V_n = ((\sigma_{ll'}))_{p\times p}=((1/(n-1)(\sum_{i=1}^{n}(Y_{1il}-Y_{2il})(Y_{1il'}-Y_{2il'}))))_{p\times p}$$

und $V_n^{-1}=((\sigma^{*}_{ll'}))_{p\times p}$.

Arnold schlägt ebenfalls keinen exakten Permutationstests vor, sondern berechnet die ersten vier Momente der Permutationsverteilung der zu T^2 in monotoner Beziehung stehenden Teststatistik $\tilde{T}^2=T^2/((n-1)+T^2)$. Royen (1978) betrachtet im Gegensatz zu Arnold den exakten zweiseitigen Permutationstest unter den 2^n zulässigen Vertauschungen der p-dimensionalen Vektoren Y_{1i} und Y_{2i}.

(iii) Während in obigen beiden Beispielen eine möglichst große Übereinstimmung mit dem parametrischen Fall in der Wahl der Teststatistik angestrebt wurde, sind für Chung & Fraser (1958) Sensitivität der gewählten Teststatistik gegen die zu erwartende Alternative und einfache Berechenbarkeit der Teststatistik für die Durchführung des exakten Permutationstests wichtiger. Folglich wird als Teststatistik für den p-dimensionalen Fall - wie auch von Edgington (1980, Kap. 7.3) vorgeschlagen - die Summe der entsprechenden eindimensionalen Teststatistiken gewählt. Explizit wird von ihnen das unverbundene Zweistichprobenproblem behandelt, aber auch betont, daß das Vorgehen trivialerweise auf jedes p-dimensionale Problem angewendet werden kann. Die von Chung & Fraser gewählten Teststatistiken sind:

$$(4.4) \qquad T = \sum_{l=1}^{p} |\bar{Y}_{1.1} - \bar{Y}_{2.1}| / \sigma_{11} \quad , \quad \sigma_{11} \text{ wie in (4.1)}$$

bzw.

$$T = \sum_{l=1}^{p} (\bar{Y}_{1.1} - \bar{Y}_{2.1})^2 / \sigma_{11}^2 \quad .$$

Falls die p abhängigen Variablen unabhängig sind, verliert man gegenüber (4.1) keine Information. Durch die Wahl von $\tilde{\sigma}_{11}$ wie in (4.2) anstelle von σ_{11} könnte man noch die wiederholte Berechnung von σ_{11} für jede der $(n_1+n_2)!/n_1!n_2!$ p-dimensionalen Vektorpermutationen bei Durchführung des exakten Permutationstests einsparen.

(iv) Boyett & Shuster (1977) halten es ähnlich wie Chung & Fraser für einen Vorteil, nicht eine quadratische Form als Teststatistik bilden zu müssen. Sie betrachten für das unverbundene wie das verbundene Zweistichprobenproblem Teststatistiken, die sensitiv gegen einseitige Alternativen

beim Vergleich zweier experimenteller Bedingungen (z.B. spezifische Therapie (i=1) vs. unspezifische Therapie (i=2) sind. Wie in (4.2) wird je abhängiger Variabler eine t-Statistik in folgender Form gebildet:

$$(4.5)\quad T_1 = ((n_1+n_2)/n_2)^{3/2}(\bar{Y}_{1.1}-\bar{Y}_{..1})/[S_1^2/n_1-((n_1+n_2)/(n_2(n_1+n_2-2))\cdot \cdot(\bar{Y}_{1.1}-\bar{Y}_{..1})^2]^{1/2}$$

mit $S_1=(n_1n_2)/(n_1+n_2)\tilde{\sigma}_{11}$ aus (4.2).

Als Teststatistik, die sensitiv gegen die spezielle Alternative

A: Bedingung 1 besser als Bedingung 2 für mindestens eine der p abhängigen Variablen ist, schlagen Boyett & Shuster vor:

$$(4.6.)\quad T_{max} = \max\{T_1,\ldots,T_p\} \quad .$$

Die Nullhypothese wird zugunsten von A verworfen, falls gilt:

$$(4.7)\quad p^{\leq}(\underline{y})=n_1!n_2!/(n_1+n_2)!\cdot|\{\pi\in\tilde{\Omega}_{n_1+n_2}\ |T_{max}(\pi\underline{y})\geq T_{max}(\underline{y})\}|\leq\alpha$$

In $\tilde{\Omega}_{n_1+n_2}$ seien alle Vertauschungen von p-dimensionalen Beobachtungen zwischen den beiden Bedingungen enthalten.

Für die Alternative

A': Bedingung 1 besser als Bedingung 2 für jede der p-abhängigen Variablen

wird dieselbe Teststatistik mit folgender gegenüber (4.7) modifizierter Entscheidungsregel benutzt:

Falls

$$(4.8)\quad p^{\leq}(\underline{y})=n_1!n_2!/(n_1+n_2)!\cdot|\{\pi\in\tilde{\Omega}_{n_1+n_2}\ |T_{max}(\pi\underline{y})\leq T_{max}(\underline{y})\}|\leq\alpha$$

ist, wird die Nullhypothese zugunsten von A' verworfen.

Für das verbundene Zweistichprobenproblem sind die

komponentenweisen Teststatistiken von $T_{max} = \max(T_1,..,T_p)$ jetzt:

(4.9) $T_l=\sqrt{n}(\bar{Y}_{1.l}-\bar{Y}_{2.l})/\{S_l^2-n/(n-1)(\bar{Y}_{1.l}-\bar{Y}_{2.l})^2\}^{1/2}$,l=1,..,p,

mit

$$S_l^2=1/(n-1)\sum_{j=1}^{n}(Y_{1jl}-Y_{2jl})^2 \quad .$$

Analog zu (4.7) und (4.8) wird unter allen 2^n zulässigen Vertauschungen der beiden p-dimensionalen Beobachtungen je Beobachtungseinheit die Testentscheidung getroffen mit $n_1!n_2!/(n_1+n_2)!$ ersetzt durch $1/2^n$.

(v) Mardia (1971) betrachtet MANOVA-Testprobleme in einem Randomisierungsmodell allgemein mit der für univariate Permutationstests üblichen Annahme der Symmetrie (d.h. $\tilde{S}_n$-Invarianz) der Verteilung der Fehler $\varepsilon_1,...,\varepsilon_n$ in den n p-dimensionalen Vektorkomponenten (vgl. Ito, 1980). Weiterhin wird (vereinfachend) vorausgesetzt, daß $rg(Var(\varepsilon_i)) = p$ sei für alle i=1,..,n. Ausführlich wird das unverbundene k-Stichprobenproblem behandelt. Mardia versucht ebenfalls, die Teststatistiken aus der parametrischen MANOVA zu benutzen. Von den verschiedenen, in der MANOVA üblicherweise verwendeten Teststatistiken hat das Spurkriterium von Pillai (1955) zur Konstruktion eines Permutationstests günstige Eigenschaften. Nach Pillai (1976) wurde es bereits von Bartlett (1939) und Nanda (1950) vorgeschlagen. Es ist gegeben durch:

(4.10) $V^{(s)} = sp(\underline{S}_h(\underline{S}_e+\underline{S}_h)^{-1})$

mit den pxp-dimensionalen 'Summen von Quadraten und Kreuzprodukten'

$$\underline{S}_h = \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})(\bar{Y}_{i.}-\bar{Y}_{..})'$$

$$\underline{S}_e = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})(Y_{ij}-\bar{Y}_{i.})' \quad .$$

$\underline{S}_e + \underline{S}_h = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij}-\bar{Y}_{..})(Y_{ij}-\bar{Y}_{..})'$ ist invariant unter den bei Gültigkeit der zur ANOVA analogen Nullhypothese zulässigen Permutationen aus $\tilde{S}_n$. Demnach muß die (verallgemeinerte) Inverse

$$((\sigma^*))_{p\times p} = S^* = (\underline{S}_e+\underline{S}_h)^{-1}$$

in der Teststatistik nur einmal berechnet werden und es gilt:

(4.11) $$V^{(s)} = \sum_{l=1}^{p} \sum_{l'=1}^{p} \left(\sum_{i=1}^{k} n_i (\bar{Y}_{i.l}-\bar{Y}_{..l})(\bar{Y}_{i.l'}-\bar{Y}_{..l'}) \right) \cdot \sigma^*_{ll'}$$

$$= \sum_{i=1}^{k} n_i (\bar{Y}_{i.}-\bar{Y}_{..})' \, S^* \, (\bar{Y}_{i.}-\bar{Y}_{..}) \quad .$$

Mardia analysiert nicht die exakte Verteilung von $V^{(s)}$ in (4.11) sondern betrachtet lediglich eine Approximation der Permutationsverteilung. Für p=2 zeigt Mikhail (1985), daß die Teststatistik von Barton & David (1962) eng mit obiger Statistik (4.11) zusammenhängt.

(vi) Der weniger eng am allgemeinen linearen Modell orientierte Ansatz der 'Multi-Response-Permutation-Procedures (MRPP)' von Mielke et al. (1976) wird ausführlicher in Abschnitt 4.7 dargestellt.

4.3 Übertragung des Konstruktionsprinzips für Invarianz-Permutationstests bezüglich linearer Hypothesen in linearen Modellen auf den multivariaten Fall

Wie in Abschnitt 4.2 erläutert, ist die Konstruktion von exakten (ähnlichen) Niveau-α Permutationstests für $p>1$ dimensionale Beobachtungen in direkter Verallgemeinerung des eindimensionalen Falles möglich (vgl. Kap. 2 und Bell & Smith, 1969).

Im parametrischen Fall ist der Übergang von der ANOVA zur MANOVA ebenfalls leicht möglich. Sowohl für den Fall fester Effekte (vgl. Anderson, 1958, S. 178, 219ff; Ahrens & Läuter, 1981; Witting, 1985, Kap. 4.4.1, S. 105ff) wie auch im Modell mit gemischten Effekten (vgl. Roebruck, 1982b) läßt sich, falls ein exakter univariater Test existiert, sofort ein multivariates Analogon angeben (s. auch Bay, 1971; Woodward & Overall, 1975). Die Summen quadrierter Abweichungen in Zähler und Nenner der F-Statistik im univariaten Fall werden durch dyadische Produkte von p-dimensionalen Vektoren derselben unquadrierten Abweichungen in den p abhängigen Variablen ersetzt (Morrison, 1976, S. 166). Da es im Gegensatz zum F-Test der ANOVA bis auf Spezialfälle keinen gleichmäßig besten (trennschärfsten) Test in der MANOVA gibt (vgl. Lehmann, 1959, S. 299f), sind in der statistischen Literatur verschiedene Teststatistiken vorgeschlagen worden. Diese lassen sich alle als Funktionen von Lösungen (Eigenwerten) einer Determinantengleichung in diesen beiden dyadischen Produkten darstellen. Die Eigenwerte im Normalverteilungsfall sind die Maximalinvarianten von bestimmten Transformationsgruppen, die das zur ANOVA analoge multivariate Testproblem invariant lassen. Die Transformationen stellen im wesentlichen Änderungen von Ursprung und Maßeinheiten dar, in denen die multivariaten Beobachtungen gewonnen werden (vgl. Lehmann, 1959, Kap. 7.10; Kres, 1975, Kap.0). Die verschiedenen Teststatistiken besitzen unterschiedliche Teststärke gegen verschiedene Abweichungen von der jeweiligen multivariaten Nullhypothese (Ahrens & Läuter, 1981, S. 52f; Davis, 1980, 1982; John, 1971; Mikhail, 1985; Olson, 1974, 1975, 1976; Schatzoff, 1966). Außerdem sind diese unterschiedlich robust gegen

verschiedene Verletzungen der vorausgesetzten multivariaten p-dimensionalen Normalverteiltheit und die Homogenität der Varianz-Kovarianz Matrizen der abhängigen Variablen (O'Brien, 1982; Ito, 1969, 1980; Olson, 1974, 1975, 1976; Stevens, 1979).

In Verallgemeinerung von (3.1) ist das verteilungsunabhängige multivariate lineare Modell gegeben durch:

$$Y_{n\times p} = X_{n\times k}\cdot b_{k\times p} + \varepsilon_{n\times p}$$

oder äquivalent, wenn man die p-dimensionalen Zeilenvektoren in Y, b und ε jeweils als Spaltenvektor aufreiht zu

$$Y_{pn\times 1}=(Y_{11},\ldots,Y_{1p},Y_{21},\ldots,Y_{2p},\ldots,Y_{n1},\ldots,Y_{np})'$$

sowie b und ε analog

(4.12) $$Y = (X\otimes I_p)b + \varepsilon \quad .$$

Der pn-dimensionale Zufallsvektor Y stellt den Vektor der p-variaten Beobachtungen dar. X bleibt die Designmatrix von vollem (Spalten)Rang k. Die Bildung des Kroneckerproduktes mit I_p bedeutet, daß für jede der p abhängigen Variablen dasselbe Design vorliegt. Die pn-dimensionale Verteilungsfunktion von Y werde charakterisiert durch den pn-dimensionalen Lageparameter $\theta=(X\otimes I_p)b$:

$$F_\theta(y) = F(y-\theta) \qquad \theta\in U\subset R^{pn} \quad .$$

U sei ein linearer Unterraum des R^{pn} mit Dimension pk. O.b.d.A. sei angenommen, daß die p-dimensionalen Randverteilungen $F_\theta(y_{i1},\ldots,y_{ip})$ nicht degeneriert seien (vgl. Koch, 1969, S. 492).

Für F soll weiterhin gelten:

$$F(x)=F(\pi x) \text{ für alle } \pi\in\Omega_n^*=\Omega_n^*\otimes\{1_p\} \text{ und } x\in R^{pn}$$

mit Ω^*_n (echte) Permutationsuntergruppe von S_n.

Zu univariaten linearen Hypothesen analoge multivariate lineare Hypothesen lassen sich wieder formulieren als:

$H: \theta \in U^H$, U^H linearer Unterraum von U der Dimension ph (4.13) oder als

$$H: \quad (K \otimes I_p)\theta = 0$$

mit idempotenter, symmetrischer npxnp Matrix $\tilde{K} = K \otimes I_p$ und K aus (3.5). $\tilde{K}$ ist die zur orthogonalen Projektion in den zu U^H orthogonalen Teilraum im R^{pn} gehörige Matrix.

Das Testproblem (4.13) ist analog zum univariaten Problem invariant unter der Gruppe $\tilde{G}$ von linearen Transformationen $\tilde{g} = id + \tilde{\theta}^*$ mit $\tilde{\theta}^* \in U^H$. Die zu $\tilde{K}$ gehörende orthogonale Projektion $\tilde{f}_H^* = id - \tilde{f}_H$ ist (maximal-) invariant bzgl. $\tilde{G}$. Der Beweis erfolgt wie im univariaten Fall.

Üblicherweise wird eine multivariate allgemeine lineare Hypothese so formuliert, daß auch Vergleiche zwischen den Komponenten der p-dimensionalen Beobachtungsvektoren möglich sind (vgl. z.B. Morrison, 1976, Kap. 5.5; Kres, 1975; Moosbrugger & Steyer, 1983, Kap. 3.3). Auf solche sog. Profilanalysen wird in Abschnitt 4.6 eingegangen. Sog. 'Growth-curve'-Verfahren (z.B. Geisser, 1980) werden nicht berücksichtigt.

Die Anwendung von $\tilde{K}$ auf Y ergibt:

$$\tilde{K}Y = (K \otimes I_p)Y = (K \otimes I_p)\theta + (K \otimes I_p)\varepsilon \quad .$$

Bei Gültigkeit der Nullhypothese hängt die Verteilungsfunktion von $\tilde{K}Y$ nicht von ε ab, sondern wird von der Verteilungsfunktion von $\tilde{K}$ induziert. Analog zu Abschnitt 3.1.3 definiert man:

$$\tilde{Q}_n = \{\pi^+ \in \tilde{S}_n \mid \text{es gibt } \pi \in Q_n^* \text{ mit } \pi^+ K = K\pi\} \quad .$$

Man erkennt sofort, daß $\tilde{Q}_n = Q_n \otimes \{1_p\}$ ist mit Q_n aus dem analogen univariaten Testproblem. $\tilde{Q}_n$ ist die Untergruppe von Vektorpermutationen, die bei Gültigkeit der multivariaten linearen Hypothese die Verteilungsfunktion der mit $\tilde{K}$ transformierten Beobachtungen unverändert läßt.

In der ANOVA war die F-Statistik

$$F = (1/df_h)SS_h / (1/df_e)SS_e$$

mit df_h, df_e Freiheitsgrade der Summen der quadrierten Abweichungen für die Hypothese und den Fehler eine Funktion der (Maximal-)Invarianten f_H^* (vgl. Abschnitt 3.1.2 mit einem Beispiel und Abschnitt 3.2 für den allgemeinen Fall!). Wie zu Beginn des Abschnitts erwähnt, sind die MANOVA-Teststatistiken Funktionen der $s = \min(p, df_h)$ von Null verschiedenen Lösungen der Determinantengleichung

$$(4.14) \quad |\underline{S}_h - \lambda \underline{S}_e| = 0$$

bzw. der Eigenwerte von $\underline{S}_e^{-1}\underline{S}_h$ oder $\underline{S}_h\underline{S}_e^{-1}$ (Notation wie in Kres, 1975).

$\underline{S}_h$ und $\underline{S}_e$ sind die zu den Summen quadrierter Abweichungen analog gebildeten dyadischen Produkte in den p-dimensionalen Beobachtungsvektoren. In (4.10) ist das Beispiel des pCR-k Planes schon demonstriert worden. Der Zusatz p an der Design-Bezeichnung nach Kirk soll andeuten, daß es sich um den multivariaten Fall handelt. $\underline{S}_h$ und $\underline{S}_e$ hängen ebenfalls 'nur' von $\tilde{f}_H^*$, der p-dimensionalen Verallgemeinerung von f_H^*, ab.

Dasselbe gilt für die zu (4.14) alternativ verwendete Determinantengleichung:

$$(4.15) \quad |\underline{S}_h - \theta(\underline{S}_e + \underline{S}_h)| = 0$$

bzw. die $u:=\min(p,df_h) + \min(p,df_e)-p$ von Null und Eins verschiedenen Eigenwerte von $(S_e+S_h)^{-1} \cdot S_h$ oder $S_h \cdot (S_e+S_h)^{-1}$ (vgl. Lehmann, 1959, S. 318, Problem 25). Da i.a. $df_h<df_e$ und $p<df_e$, ist i.a. u=s. Auch Funktionen der Lösungen von

(4.16) $$|\underline{S}_e - \mu(\underline{S}_e+\underline{S}_h)| = 0$$

sind betrachtet worden.

Kres (1975, Formel (7) - (9) in Kap. 0) gibt die Beziehungen zwischen den Lösungen von (4.14) - (4.16) an:

$$\lambda_i = \theta_i/(1-\theta_i) = (1-\mu_k)/\mu_k \qquad \begin{matrix} i=1,..,s \\ k=s,s-1,...,1 \end{matrix}$$

(4.17) $$\theta_i = \lambda_i/(1+\lambda_i) = 1-\mu_k$$

$$\mu_k = 1/(1+\lambda_i) = 1 - \theta_i \quad .$$

In Kres findet sich auch eine Zusammenstellung von 12 in der Literatur vorgeschlagenen Teststatistiken gebildet mit den Eigenwerten aus (4.17). Die wichtigsten sind:

(i) 'Spurkriterien' von Pillai bzw. Hotelling:

$$V^{(s)} = sp(\underline{S}_h(\underline{S}_e+\underline{S}_h)^{-1}) = \sum_{i=1}^{s} \theta_i$$

$$V = sp(\underline{S}_h\underline{S}_e^{-1}) = \sum_{i=1}^{s} \lambda_i$$

(4.18) (ii) 'Quotienten von Determinanten' von Wilks:

$$\Lambda = |\underline{S}_e|/|\underline{S}_e+\underline{S}_h| = \prod_{i=1}^{s} \mu_i$$

$$U = |\underline{S}_h|/|\underline{S}_e+\underline{S}_h| = \prod_{i=1}^{s} (1/(1-\theta_i))$$

(iii) 'Maximalwurzelkriterien' von Roy

$$\lambda_{max} \quad \text{bzw.} \quad \theta_{max}$$

Eine Darstellung der Eigenwerte als Quadrate kanonischer Korrelationen zwischen unabhängigen und abhängigen Variablen ist in Bredenkamp (1980, S. 81f) und Bredenkamp & Erdfelder (1985) zu finden.

Für ein (heuristisches) Konstruktionsprinzip für multivariate Invarianzpermutationstests in Multiparametermodellen liegt es somit nahe, die Permutations-Invarianzeigenschaften der Verteilungsfunktion der um störende Effekte bereinigten p-dimensionalen Beobachtungen zu nutzen. Als Teststatistiken sind die Teststatistiken der MANOVA geeignet. In diesen sind lediglich die ursprünglichen Zufallsvektoren Y_i durch ihre Transformierten $T_i=(\tilde{KY})_i$ zu ersetzen.

Bei der Konstruktion univariater Invarianz-Permutationstests hatte sich ergeben, daß die verwendeten Teststatistiken äquivalent zu den Summen quadrierter Abweichungen in den Zählern der F-Statistiken sind, wenn dort die ursprünglichen durch die transformierten Beobachtungen ersetzt werden. Wenn man zur Herleitung multivariater Permutationstests diesen Ansatz aufgreift, müßte man Funktionen lediglich von $\underline{S}_h(Y)$ bzw. $\underline{S}_h(KY)$ verwenden. Bei Bildung der Spur von $\underline{S}_h$ erhält man dann die Summe der univariaten Statistiken. Testentscheidungen bzgl. $sp(\underline{S}_h)$ sind offensichtlich aber nicht invariant unter verschiedenen linearen Transformationen der p abhängigen Variablen. Deshalb müssen Funktionen der Eigenwerte aus (4.14) oder (4.15) gewählt werden.

4.4 Beispiel: Test auf Wechselwirkungen im multivariaten Split-Plot Plan (pSPF-r.c, Willmes, 1981a)

In Verallgemeinerung von Abschnitt 3.1.5 (xi) wird im dort beschriebenen Fall (a) angenommen, daß je Beobachtungseinheit die Werte von p>1 abhängigen Variablen erhoben werden. Es soll ein Test auf Wechselwirkungen zwischen den beiden experimentellen Faktoren A (Plot) und B (Split)

konstruiert werden.

Das Modell für die p-dimensionalen Zufallsvariablen lautet in Analogie zu (3.61):

$$Y_{ih(i)j} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\pi_{h(i)}+\varepsilon_{ih(i)j} \qquad \begin{matrix} i=1,..,r \\ h(i)=1,..,n_i \\ n_.=\sum n_i \\ j=1,..,c \\ N=n_.c \end{matrix}$$

(4.19) mit $\alpha_.=\beta_.=(\alpha\beta)_{i.}=(\alpha\beta)_{.j}=\pi_{.(i)} = 0$.

Der gesamte Fehlervektor

$\varepsilon = (\varepsilon'_{111},\cdots,\varepsilon'_{rn_rc})'$ mit $\varepsilon_{ih(i)j}=(\varepsilon_{ih(i)j1},\cdots,\varepsilon_{ih(i)jp})'$

sei verteilt nach $F(x)$, $x\in R^{pN}$, mit

$F(x)=F(\pi x)$ für alle $\pi\in\mathcal{Q}_n^*=S_c\wr S_{n_.}=S_c\otimes\{1_p\}\wr S_{n_.}\otimes\{1_p\}$;

d.h. die pc-dimensionalen Fehler $\varepsilon_{ih(i)}=(\varepsilon'_{ih1},\cdots,\varepsilon'_{ihc})'$ haben vertauschbare p-dimensionale Vektorkomponenten; außerdem seien sie über alle n. Blöcke vertauschbar.

Damit ist vorausgesetzt, daß je Block die Matrizen $\mathrm{Cov}(\varepsilon_{ihj},\varepsilon_{ihj'})$ identisch sind für $j,j' = 1,..,c$ und die Blöcke keine Variabilitätsunterschiede für verschiedene Stufen des Plotfaktors A zeigen dürfen. Ansonsten wäre die Vertauschbarkeit über die Stufen des Plotfaktors hinweg verletzt.

Getestet werden soll:

$$H: (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{rc} = 0$$

(4.20) gegen

$$A: (\alpha\beta)_{ij} \neq (\alpha\beta)_{i'j'} \quad \text{für ein } (i,j)\neq(i',j') \qquad i,i'=1,..,r;\ j,j'=1,..,c.$$

In direkter Verallgemeinerung des univariaten Falles hat man (zur Vereinfachung der Schreibweise wird die Abhängigkeit des Index h(i) vernachlässigt):

$$U = \{\theta \in R^{pN} \mid \theta_{ihj} = \mu + \alpha_i + \beta_j + \pi_h + (\alpha\beta)_{ij};\ \mu, \alpha_i, \beta_j, \pi_h, (\alpha\beta)_{ij} \in R^p\}$$

$$U^H = \{\theta \in U \mid \theta_{ihj} = \mu + \alpha_i + \beta_j + \pi_h\}$$

$$\theta \in U^H \iff \theta_{ihjl} - \bar{\theta}_{ih.l} - \bar{\theta}_{..jl} + \bar{\theta}_{...l} = 0 \quad \text{für alle } l = 1,..,p \ .$$

Damit ist $\tilde{K} = K \otimes I_p$

(4.21) $= ((I_{n.} - 1/n_. J_{n.}) \otimes (I_c - 1/c J_c)) \otimes I_p$

und die transformierten Beobachtungen sind:

(4.22) $(\tilde{K}Y)_{ihjl} =: T_{ihjl} = Y_{ihjl} - \bar{Y}_{ih.l} - \bar{Y}_{..jl} + \bar{Y}_{...l}$.

Ebenso sieht man sofort:

(4.23) $\tilde{\mathcal{Q}}_N = \tilde{S}_N(\tilde{K}) = \tilde{S}_{n.} \otimes \tilde{S}_c = (S_{n.} \otimes \{1_p\}) \otimes (S_c \otimes \{1_p\})$.

Zur Bestimmung einer Teststatistik werden $\underline{S}_h$ und $\underline{S}_e$ des MANOVA-Problems benutzt:

$$\underline{S}_h(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_i (\bar{Y}_{i.j} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})(\bar{Y}_{i.j} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})'$$

$$\underline{S}_e(Y) = \sum_{i=1}^{r} \sum_{h=1}^{n_i} \sum_{j=1}^{c} (Y_{ihj} - \bar{Y}_{i.j} - \bar{Y}_{ih.} + \bar{Y}_{i..})(Y_{ihj} - \bar{Y}_{i.j} - \bar{Y}_{ih.} + \bar{Y}_{i..})'$$

$$(\underline{S}_e + \underline{S}_h)(Y) = \sum_{i=1}^{r} \sum_{h=1}^{n_i} \sum_{j=1}^{c} (Y_{ihj} - \bar{Y}_{ih.} - \bar{Y}_{..j} + \bar{Y}_{...})(Y_{ihj} - \bar{Y}_{ih.} - \bar{Y}_{..j} + \bar{Y}_{...})'$$

Mit den Beziehungen $\bar{T}_{i.j} = \bar{Y}_{i.j} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...}$

und $\bar{T}_{ih.} = \bar{T}_{i..} = \bar{T}_{..j} = \bar{T}_{...} = 0$ hat man:

$$\underline{S}_h(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_i \bar{T}_{i.j} \bar{T}'_{i.j}$$

(4.24) $$\underline{S}_e(T) = \sum_{i=1}^{r} \sum_{h=1}^{n_i} \sum_{j=1}^{c} (T_{ihj} - \bar{T}_{i.j})(T_{ihj} - \bar{T}_{i.j})'$$

$(\underline{S}_e + \underline{S}_h)(T) = \sum\sum\sum T_{ihj} T'_{ihj}$.

$(\underline{S}_e + \underline{S}_h)(T)$ wie auch der erste Term von $\underline{S}_e$ sind invariant unter den zulässigen Permutationen aus $\tilde{\mathcal{Q}}_n$. Für diesen Versuchsplan wie auch allgemein gilt: Unter der Nullhypothese sind alle Summanden, die von Y_{ihjl} in (4.22) subtrahiert werden, permutationsinvariant unter $\tilde{\mathcal{Q}}_N$. Sie stellen ja komponentenweise die Projektion f_H in den Hypothesenunterraum

dar. Da $\tilde{f}_H^* = id - \tilde{f}_H$ und $\tilde{K}Y$ $\tilde{\Omega}_N$-invariant ist unter der Nullhypothese, ist $\tilde{K}_H Y$ mit $\tilde{K}_H$ die zu $\tilde{f}_H$ gehörige Matrix - ebenfalls trivialerweise $\tilde{\Omega}_N$-invariant. Maximal $n_\cdot!/\prod_{i=1}^{r} n_i!$ von diesen Permutationen führen zu verschiedenen Werten von $\underline{S}_h$ bzw. $\underline{S}_e$. Die Vertauschungen aus $\tilde{S}_C$ bewirken ebenfalls nur Umindizierungen in $\underline{S}_h$ sowie $\underline{S}_e$.

Bei der Wahl einer Teststatistik ist für einen Permutationstest im Gegensatz zu den MANOVA-Verfahren auch zu berücksichtigen, welche algebraischen Operationen bei der praktischen Durchführung des Tests wie oft auszuführen sind. In (4.18) sind zum einen Determinanten- und Eigenwertberechnungen viel komplexer als die Bildung einer Spur. Zum anderen sind die Kriterien weniger aufwendig, die $\underline{S}_h + \underline{S}_e$ bzw. die Inverse davon enthalten. Beide müssen je Testdurchführung wegen ihrer Permutationsinvarianz nur einmal bestimmt werden. Kriterien mit $\underline{S}_e^{-1}$ erfordern die Berechnung einer Inversen für jede zulässige Permutation. Unter diesen beiden Aspekten hat das Spurkriterium $V^{(s)}$ von Pillai günstige Eigenschaften. Man findet mit $(S_e+S_h)^- := ((\sigma^*_{ll'}))_{p\times p} = S^*$ als (verallgemeinerter) Inverser:

$$V^{(s)}(T) = sp(\underline{S}_h(T)\cdot S^*) \tag{4.25}$$
$$= \sum_{l=1}^{p}\sum_{l'=1}^{p} \sigma^*_{ll'} \sum_{i=1}^{r}\sum_{j=1}^{c} n_i \bar{T}_{i.jl}\cdot\bar{T}_{i.jl'} \quad .$$

Mit der bekannten Abzählregel für den im Experiment realisierten pN-dimensionalen Beobachtungsvektor t:

$$p^{>} = |\{V^{(s)}(\pi t) > V^{(s)}(t),\ \pi\in\tilde{\Omega}_N\}|/(n_\cdot!/\prod_{i=1}^{r} n_i!)$$
$$p^{=} = |\{V^{(s)}(\pi t) = V^{(s)}(t),\ \pi\in\tilde{\Omega}_N\}|/(n_\cdot!/\prod_{i} n_i!) \tag{4.26}$$

wird die Nullhypothese (4.20) für $0<\alpha<1$ verworfen, falls $p := p^{>} + p^{=} \leq \alpha$ ist. Bei anderen multivariaten Teststatistiken,

wie z.B. Wilks' λ ist in (4.26) $>$ durch $<$ zu ersetzen, da hier 'sehr kleine' Werte der Teststatistik gegen die Nullhypothese sprechen.

Auch der Fall (b) in Abschnitt 3.1.5 (xi) ist direkt auf multivariate Beobachtungen verallgemeinerbar. Beispielsweise könnte der Split-Faktor Zeitpunkte und die p-variaten Beobachtungen Werte in einem p-variaten Testprofil sein.

Man hat als Modell:

$$Y_{ih(i)jl} = \mu_{ih(i)jl} + \varepsilon_{ih(i)jl} \qquad \begin{matrix} i,h(i),j \text{ wie zuvor,} \\ l=1,..,p \end{matrix}$$

mit den $n_.$ pc-dimensionalen Profilen

$$\mu_{ih(i)} = (\mu'_{ih(i)1},\ldots,\mu'_{ih(i)c})'$$

und den uber alle n. Blocke vertauschbaren pc-dimensionalen 'Fehler-Profilen', d.h.:

$$F(x)=F(\pi x) \quad \text{für alle } \pi \in \tilde{S}_{n_.} = S_{n_.} \otimes \{1_{pc}\} \ .$$

Die Hypothese der Profilparallelität bzgl. des Split-Faktors läßt sich analog zu (3.63) darstellen als:

$$H: \text{ Es gibt } (\Delta'_1,..,\Delta'_{c-1})' \in R^{p(c-1)}; \ \Delta'_j=(\Delta_{j1},..,\Delta_{jp})$$

mit

$$\mu_{ih(i)j} - \mu_{ih(i),j+1} = \Delta_j \qquad j=1,..,c-1$$

oder äquivalent:

$$H: \ \mu_{ih(i)j} - \bar{\mu}_{ih(i).} - \bar{\mu}_{..j} + \bar{\mu}_{...} = 0$$

$$\text{d.h. } \tilde{K}\mu = 0 \quad \text{mit } \tilde{K} \text{ wie in (4.21)}$$

gegen

$$A: \quad \tilde{K}\mu \neq 0 \ .$$

Hier ist $\tilde{\Sigma}_{n_.} = S_{n_.} \otimes \{1_{pc}\}$ und $\underline{S}_h$ und $\underline{S}_e$ werden genau wie zuvor gebildet.

Wie im univariaten Fall ist auch hier alternativ die Bildung einer quadratischen Form in den p-dimensionalen $\bar{T}_{i.j}$ möglich.

Für die p bzw. 2p-dimensionalen Randverteilungen von $T = (T'_{111},\ldots,T'_{rcn_r})'$ gilt bei gegebenem t in Analogie zum univariaten Fall:

$$\tilde{P}_N(T_{ihj}=x) = 1/n_{\cdot}c \quad , x\in R^p$$

$$\tilde{P}_N(T_{ihj}=x,\ T_{i'h'j'}=x') = \begin{cases} 1/(n_{\cdot}c(c-1)) & (i,h)=(i',h'),\ j\neq j' \\ 1/(n_{\cdot}(n_{\cdot}-1)c & (i,h)\neq(i',h'),\ j=j' \\ 1/(n_{\cdot}(n_{\cdot}-1)c(c-1)) & (i,h)\neq(i',h'),\ j\neq j' \end{cases}$$

$$E_{\tilde{P}_N}(T_{ihj}) = t_{\ldots} = 0 \ ,$$

$$Var_{\tilde{P}_N}(T_{ihj}) =: ((n_{\cdot}-1)(c-1))/n_{\cdot}c\cdot\Sigma_{N,t} \quad \text{mit } \Sigma_{N,t}=((\sigma^{ll'}_{N,t}))_{p\times p} \text{ und}$$

$$\sigma^{ll'}_{N,t}=1/((n_{\cdot}-1)(c-1))\cdot\sum_{ihj}\sum\sum t_{ihjl}\cdot t_{ihjl'}$$

$$Cov_{\tilde{P}_N}(T_{ihj},T_{ihj'}) = -(n_{\cdot}-1)/n_{\cdot}c\cdot\Sigma_{N,t}$$

$$Cov_{\tilde{P}_N}(T_{ihj},T_{i'h'j}) = -(c-1)/n_{\cdot}c\cdot\Sigma_{N,t}$$

$$Cov_{\tilde{P}_N}(T_{ihj},T_{i'h'j'}) = 1/n_{\cdot}c\cdot\Sigma_{N,t} \quad .$$

Für die $\bar{T}_{i.j}$ findet man in Analogie zum univariaten Fall dieselben Koeffizienten für die pxp-dimensionale Varianz-Kovarianz Matrix $\Sigma_{N,t}$ wie dort von $\sigma^2_{N,t}$.

Mit $T_N=(\bar{T}'_{1.1},..,\bar{T}'_{r.c})'$ hat man

$$E_{\tilde{P}_N}(T_N) = 0 \quad \text{und}$$

$$Var_{\tilde{P}_N}(T_N) = (((n_{\cdot}/n_i\delta_{ii'}-1))\otimes(I_c-1/cJ_c))\otimes 1/n_{\cdot}\Sigma_{N,t}$$

mit $(r-1)(c-1)\leq rg(V_{\tilde{P}_N}(T_N))\leq(r-1)(c-1)p$.

Es sei im folgenden immer o.B.d.A. angenommen, daß die p-dimensionalen Vektoren T_{ihj} eine Varianz-Kovarianz Matrix von vollem Rang p besitzen.

Als verallgemeinerte Inverse von $V_{\tilde{P}_N}(T_N)$ hat man:

$$V_N^{-} = n_{\cdot}((n_i/n_{\cdot}\delta_{ii'} \pm n_in_{i'}/n_r))\otimes(I_c-1/cJ_c)\otimes\Sigma^{-}_{N,t}$$

und als quadratische Form:

(4.27) $Q(T_N) = \sum_{l=1}^{p}\sum_{l'=1}^{p} \sigma^{*ll'}_{N,t} \sum_{i=1}^{r}\sum_{j=1}^{c} n_i \bar{T}_{i.jl}\cdot\bar{T}_{i.jl'} = (n_.-1)(c-1)V^{(s)}$

wegen (4.25) .

Im Fall (b) findet man:

$\tilde{P}_{n_.}(T_{ihj} = x) = 1/n.$

$$\hat{P}_{n_.}(T_{ihj}=x, T_{i'h'j'}=x') = \begin{cases} 1/n_. & (i,h)=(i',h') \ , \ j\neq j' \\ 1/(n_.(n_.-1) & (i,h)\neq(i',h') \ , \ j,j' \text{ beliebi} \end{cases}$$

$E_{\hat{P}_{n_.}}(T_{ihj}) = \bar{t}_{..j} = 0$

$\mathrm{Var}_{\hat{P}_{n_.}}(T_{ihj}) =: (n_.-1)/n_.\cdot\Sigma_{jj}$ mit $\Sigma_{jj}=((\sigma^{ll'}_{jj}))_{p\times p}$ und

$$\sigma^{ll'}_{jj} = 1/(n_.-1)\sum_{i=1}^{r}\sum_{h=1}^{n_i} t_{ihjl}\cdot t_{ihjl'}$$

$\mathrm{Cov}_{\hat{P}_{n_.}}(T_{ihj},T_{ihj'}) =: (n_.-1)/n_.\cdot\Sigma_{jj'}$ mit $\Sigma_{jj'}=((\sigma^{ll'}_{jj'}))_{p\times p}$ und

$$\sigma^{ll'}_{jj'} = 1/(n_.-1)\sum_{i=1}^{r}\sum_{h=1}^{n_i} t_{ihjl}\cdot t_{ihj'l'}$$

$\mathrm{Cov}_{\hat{P}_{n_.}}(T_{ihj},T_{i'h'j}) = -1/n_.\Sigma_{jj}$; $\mathrm{Cov}_{\hat{P}_{n_.}}(T_{ihj},T_{i'h'j'}) = -1/n_.\Sigma_{jj'}$.

Mit denselben Methoden wie im univariaten Fall erhält man:

$E_{\hat{P}_{n_.}}(T_N) = 0$ und

$V_{\hat{P}_{n_.}}(T_N) = 1/n_.\cdot((n_./n_i\delta_{ii'}-1)\otimes\hat{\Sigma}_{n_.}$ und $\hat{\Sigma}_{n_.}=((\Sigma_{jj'}))_{pc\times pc}$.

Damit ist eine verallgemeinerte Inverse gegeben durch:

$$V^{-}_{n_.} = n_.((n_i/n_.\delta_{ii'} \pm n_i n_{i'}/n_r))\otimes\hat{\Sigma}^{-}_{n_.} \ ,$$

und als Teststatistik ergibt sich die quadratische Form:

(4.28) $$Q(T_N) = \sum_{i=1}^{r} n_i \sum_{j=1}^{c}\sum_{j'=1}^{c}\sum_{l=1}^{p}\sum_{l'=1}^{p} \bar{T}_{i.jl}\sigma^{*ll'}_{jj'}\bar{T}_{i.j'l'} \ .$$

Asymptotisch besitzt Q eine χ^2_f-Verteilung mit $(r-1)p \leq f \leq (r-1)(c-1)p$ Freiheitsgraden.

4.5 Multivariate Invarianz-Permutationstests für ein- bis dreifaktorielle Versuchspläne

Mit dem Konstruktionsprinzip aus Abschnitt 4.3 können in Verallgemeinerung von Abschnitt 3.1.5 die Tests für alle dort betrachteten univariaten Versuchspläne auf den multivariaten Fall übertragen werden. Die Bezeichnungen für den univariaten Fall werden übernommen. Die eindimensionalen Zufallsvariablen dort sind lediglich stets durch $p>1$-dimensionale Zufallsvektoren zu ersetzen. Die Permutationsgruppen $\mathcal{G}_n^*$ bzw. $\mathcal{G}_n$, unter denen die Verteilungsfunktionen der Fehler bzw. die Verteilungsfunktion der transformierten Beobachtungen bei Gültigkeit der jeweiligen Nullhypothese invariant sind, bestehen nun aus Vektorpermutationen, bei denen stets die p-dimensionalen Beobachtungsvektoren vertauscht werden. Sie sind also isomorph zu den Permutationsgruppen der analogen univariaten Testprobleme: $\mathcal{G}_n^* = \mathcal{Q}_n^* \otimes \{\iota_p\}$ bzw. $\mathcal{G}_n = \mathcal{Q}_n \otimes \{\iota_p\}$.

Nachfolgend werden für die zu testende(n) Hypothese(n) in jedem Versuchsplan die Matrizen $\underline{S}_h$, $\underline{S}_e$ und $\underline{S}_h + \underline{S}_e$ in den um die störenden Effekte bereinigten Beobachtungen benötigt. Aus diesen können dann die verschiedenen Teststatistiken für die multivariaten Invarianz-Permutationstests wie in der MANOVA gebildet werden. Zur Durchführung der Permutationstests werden wie üblich die die Teststatistik nicht invariant lassenden Permutationen aus $\mathcal{G}_n$ ausgeführt und $p^>$ (bzw. $p^<$ je nach Testkriterium) sowie $p^=$ bestimmt. Die Testentscheidung wird dann wie in Abschnitt 4.4 beschrieben getroffen.

(i) Unverbundenes Zweistichprobenproblem (Wald & Wolfowitz, 1944)

In Analogie zu (3.37) wird für den zweiseitigen Test (4.2) als Teststatistik verwendet. Tests, die sensitiv gegen spezielle einseitige Alternativen sind, wurden bereits in Abschnitt 4.2 (iv) angeführt (Boyett & Shuster, 1977).

(ii) Verbundenes Zweistichprobenproblem (Arnold, 1964)

In Analogie zu (3.39) wird für den zweiseitigen Test (4.3) als Teststatistik gewählt. Für spezielle einseitige Alternativen können wiederum die Teststatistiken von Boyett & Shuster (1977) benutzt werden (vgl. Abschnitt 4.2 (iv)!).

(iii) Unverbundenes k-Stichprobenproblem (pCR-k, Mardia, 1971; Willmes, 1981a, b)

Da keine störenden Effekte vorhanden sind, können wie in (4.10) bereits angegeben, die nicht transformierten Beobachtungsvektoren $Y_{ij}=(Y_{ij1},\ldots,Y_{ijp})'$ zur Bildung von

(4.29) $$\underline{S}_h(Y)=\sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})(\bar{Y}_{i.}-\bar{Y}_{..})' \; ; \; \underline{S}_e(Y)=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})(Y_{ij}-\bar{Y}$$

$$(\underline{S}_h+\underline{S}_e)(Y)=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{..})(Y_{ij}-\bar{Y}_{..})'$$

verwendet werden.

Zulässig bei Gültigkeit der Nullhypothese sind alle $n.!$ Vektorpermutationen $\tilde{Q}_{n.} = \tilde{S}_{n.}$. Diese lassen $\underline{S}_h + \underline{S}_e$ und den zweiten Term in $\underline{S}_h(Y)=\sum_i n_i\bar{Y}_{i.}\bar{Y}_{i.}' - n_{..}\bar{Y}_{..}\bar{Y}_{..}'$ invariant. Nur die $n_{.}!/\prod_{i=1}^{k} n_i!$ Vektorpermutationen zwischen Beobachtungen aus verschiedenen der k Stichproben können zu

unterschiedlichen Werten einer der MANOVA analogen Teststatistiken führen, denn auch der erste Term in $\underline{S}_e(Y) = \sum_{ij}\sum Y_{ij}Y'_{ij} - \sum_i n_i \bar{Y}_{i.}\bar{Y}'_{i.}$ ist permutationsinvariant.

(iv) Verbundenes k-Stichprobenproblem (pRB-k; Willmes, 1981a)

In Verallgemeinerung der Ergebnisse aus Abschnitt 3.1.2 hat man für die Beobachtungsvektoren $Y_{ij} = (Y_{ij1},..,Y_{ijp})'$ - i=1,..,r Blöcke, j=1,..,k Faktorstufen, n=kr- und den Test auf Effekte des k-stufigen Faktors:

Die Matrix zur Eliminierung der störenden Blockeffekte ist $\overset{\circ}{K} = K \otimes I_p = (I_{rk} - 1/kI_r \otimes J_k) \otimes I_p$ (vgl. 3.10) und die transformierten p-dimensionalen Beobachtungsvektoren:

(4.30) $T_{ij} = (\overset{\circ}{K}Y)_{ij} = Y_{ij} - \bar{Y}_{i.}$.

Zulässige Permutationen bei Gültigkeit der Nullhypothese sind die $(k!)^r$ Intrablockpermutationen der p-dimensionalen T_{ij} und r! Vertauschungen ganzer Blöcke. Damit ist $\overset{\circ}{\Omega}_n = \overset{\circ}{S}_k \wr \overset{\circ}{S}_r$. Einsetzen von T_{ij} in die analog zu SS_h und SS_e (vgl. Kirk, 1968, S. 131f) gebildeten dyadischen Produkte $\underline{S}_h$ und $\underline{S}_e$ ergibt mit $\bar{T}_{i.} = \bar{T}_{..} = 0$:

(4.31) $$\underline{S}_h(T) = \sum_{j=1}^{k} r\bar{T}_{.j}\bar{T}'_{.j} \; ; \; \underline{S}_e(T) = \sum_{i=1}^{r}\sum_{j=1}^{k} (T_{ij} - \bar{T}_{.j})(T_{ij} - \bar{T}_{.j})'$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{i=1}^{r}\sum_{j=1}^{k} T_{ij}T'_{ij} \; ; \; = \sum_{i=1}^{r}\sum_{j=1}^{k} T_{ij}T'_{ij} - \sum_{j=1}^{k} r\bar{T}_{.j}\bar{T}'_{.j} \; .$$

$\underline{S}_h + \underline{S}_e$ und der erste Term in $\underline{S}_e$ sind invariant unter $\tilde{\Omega}_n$. $\underline{S}_h$ und $\underline{S}_e$ sind ebenfalls invariant unter den Vertauschungen der ganzen Blöcke, sodaß lediglich die $(k!)^r$ Intrablockpermutationen ausgeführt werden müssen. Für die praktische Durchführung des Permutationstests ist noch von Interesse, daß in T_{ij} die Blockmittel $\bar{Y}_{i.}$ permutationsinvariant sind und damit

für die Durchführung des Tests nur einmal in allen r Blöcken berechnet werden müssen.

(v) Verbundenes k-Stichprobenproblem mit Wiederholungen (pRB-kmW)

Die p-dimensionalen Beobachtungsvektoren seien $Y_{ijl}, i=1,..,r$ Blöcke, $j=1,..,n_{il}>1$, $l=1,..,k$ Faktorstufen, $n_{..} = \sum_{i=1}^{r} \sum_{l=1}^{k} n_{il}$.
Die transformierten Zufallsvektoren sind dann mit K aus (3.43) und $\mathring{K} = \bigoplus_{i=1}^{r} ((I_{n_{i.}} - 1/n_{i.} J_{n_{i.}}) \otimes I_p)$:

(4.32) $T_{ijl} = Y_{ijl} - \bar{Y}_{i..}$.

Bei Gültigkeit der Nullhypothese wie in (iv) sind die $\prod_{i=1}^{r} n_{i.}!$ Intrablockpermutationen der T_{ijl} zulässig:

$\mathring{Q}_{n..} = \bigtimes_{i=1}^{r} \mathring{S}_{n_{i.}}$. Mit $\bar{T}_{i..} = \bar{T}_{...} = 0$ gilt:

(4.33) $\underline{S}_h(T) = \sum_{l=1}^{k} n_{.1} \bar{T}_{..1} \bar{T}'_{..1}$; $\underline{S}_e(T) = \sum\sum\sum_{ijl} (T_{ijl} - \bar{T}_{..1})(T_{ijl} - \bar{T}_{..1})'$

$(\underline{S}_e + \underline{S}_h)(T) = \sum\sum\sum_{ijl} T_{ijl} T'_{ijl}$ $= \sum\sum\sum_{ijl} T_{ijl} T'_{ijl} - \sum_{l} n_{.1} \bar{T}_{..1} \bar{T}'_{..1}$

$\underline{S}_h$ + $\underline{S}_e$ sowie der erste Term in $\underline{S}_e$ sind invariant unter $\mathring{Q}_{n..}$; dasselbe gilt für die Blockmittel $\bar{Y}_{i..}$ aus (4.32).

(vi) Ausgewogener unvollständiger Blockplan (pBIB)

Mit der Schreibweise aus (3.45) hat man als Beobachtungsvektoren Y_{ij} mit $(i,j) \in D = \{L_{ij} | L_{ij}=1\}$ und $L_{ij} = 1$, falls Faktorstufe j in Block i auftritt.
Die transformierten Zufallsvektoren sind:

(4.34) $T_{ij} = Y_{ij} - 1/c \sum_{j'=1}^{k} L_{ij'} Y_{ij'}$, für $(i,j) \in D$ und c Faktorstufen je Block.

Bei Gültigkeit der Nullhypothese aus (iv) sind die $(c!)^r$ Intrablockpermutationen $\tilde{\mathcal{G}}_n = \overset{r}{\underset{i=1}{\times}} \mathcal{S}_{L_i}$ der T_{ij} zulässig. Mit $\bar{T}_{i.} = 1/c \sum_{j'=1}^{k} L_{ij'} T_{ij'} = 0 = \bar{T}_{..}$ findet man wieder (vgl. Kirk, 1968, S. 428,431; Hicks, 1973, S. 67-69):

$$\underline{S}_h(T) = cs^2/k\lambda \sum_{j=1}^{k} \bar{T}_{.j}\bar{T}'_{.j} \quad \text{mit } \bar{T}_{.j} = 1/s \sum_{i=1}^{r} L_{ij} T_{ij}$$

$$\underline{S}_e(T) = \sum_{i=1}^{r} \sum_{j=1}^{k} L_{ij} T_{ij} T'_{ij} - cs^2/k\lambda \sum_{j=1}^{k} \bar{T}_{.j}\bar{T}'_{.j} \tag{4.35}$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{i=1}^{r} \sum_{j=1}^{k} L_{ij} T_{ij} T'_{ij} \quad .$$

$\underline{S}_h + \underline{S}_e$ sowie der erste Teil von $\underline{S}_e$ sind invariant unter $\tilde{\mathcal{G}}_n$ wie auch $1/c \sum_{j'=1}^{k} L_{ij'} Y_{ij'}$ für alle i=1,..,n.

(vii) Einfach hierarchischer zweifaktorieller Zufallsplan (pCRH-r(c))

Die p-dimensionalen Beobachtungen seien

$$Y_{ij(i)k} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ij(i)k}$$

i=1,..,r Stufen Faktor A,
j(i)=1,..,c(i) Stufen von Faktor B genestet unter Stufe A_i,
k=1,..,$n_{ij(i)}$ Replikationen;

$n_{i.} = \sum_{j(i)=1}^{c(i)} n_{ij(i)}$; $n = \sum_{i} \sum_{j(i)} n_{ij(i)}$ mit innerhalb der Faktorstufen von A vertauschbaren Fehlerkomponenten $\mathcal{G}_n^* = \overset{r}{\underset{i=1}{\times}} \mathcal{S}_{n_{i.}}$. Um zu testen, ob der genestete Faktor B einen Einfluß hat, müssen die Beobachtungen um die Haupteffekte des übergeordneten Faktors A mit $K = \overset{r}{\underset{i=1}{\oplus}} ((I_{n_{i.}} - 1/n_{i.} J_{n_{i.}}) \otimes I_p)$ (vgl. 3.1.5 (vii)) bereinigt werden:

$$T_{ij(i)k} = Y_{ij(i)k} - \bar{Y}_{i..} \quad . \tag{4.36}$$

Bei Gültigkeit der Nullhypothese 'keine Effekte des genesteten Faktors B' ist dann $\mathcal{G}_n = \overset{r}{\underset{i=1}{\times}} \mathcal{S}_{n_{i.}}$; d.h. alle Permutationen innerhalb der Stufen von A sind zulässig. In Analogie zum

univariaten Plan hat man:

(4.37) $\underline{S}_h(T) = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} n_{ij} \bar{T}_{ij(i)} . \bar{T}'_{ij(i)}$. und weiterhin

(vgl. Kirk, 1968, S. 232f)

$$\underline{S}_e(T) = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} \sum_{k=1}^{n_{ij}} T_{ij(i)k} T'_{ij(i)k} - \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} n_{ij} \bar{T}_{ij(i)} . \bar{T}'_{ij(i)}$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{i=1}^{r} \sum_{j(i)=1}^{c(i)} \sum_{k=1}^{n_{ij}} T_{ij(i)k} T'_{ij(i)k} \quad .$$

$\underline{S}_h$ + $\underline{S}_e$ wie auch der erste Teil von $\underline{S}_e$ sind invariant unter $\tilde{\Omega}$ wie auch $\bar{Y}_{i..}$ für alle i aus (4.36). Nur die $\prod_{i=1}^{r}(n_{i.}! / \prod_{j(i)=1}^{c(i)} n_{ij}!)$ Permutationen zwischen verschiedenen Stufen des genesteten Faktors B können zu unterschiedlichen Werten der Teststatistik führen.

(viii) Weitere hierarchische - und partiell hierarchische Versuchspläne

In direkter Verallgemeinerung von (viii) und dem üblichen Übergang zu p-dimensionalen Beobachtungsvektoren gelten die Ergebnisse von Abschnitt 3.1.5 (vii) auch hier. Deshalb werden die Resultate ohen weitere Erläuterungen mit den Bezeichnungen analog zum univariaten Fall angeführt.

pCRH-r(c)(s): (vgl. Kirk, 1968, S. 234f) als Beobachtungen hat man $Y_{ij(i)k(i,j)l}$ mit $\tilde{\Omega}^{*}_{n.} = \times_{i=1}^{r} \times_{j(i)=1}^{c(i)} \tilde{S}_{n_{ij}}$ - invarianten Fehlern.

Zum Testen des Effektes des in der Hierarchie niedrigsten genesteten Faktors C werden die Beobachtungen mit

$$\tilde{K} = \bigoplus_{i=1}^{r} \bigoplus_{j(i)=1}^{c(i)} ((I_{n_{ij}} - 1/n_{ij} . J_{n_{ij}}) \otimes I_p)$$

transformiert zu:

(4.38) $T_{ij,(i)k(i,j)l} = Y_{ij(i)k(i,j)l} - \bar{Y}_{ij(i)..}$.

Bei Gültigkeit der Nullhypothese 'kein Effekt von Faktor C' ist $\hat{Q}_n = \hat{Q}^*_n$ und:

(4.39)
$$\underline{S}_h(T) = \sum_{ijk} n_{ijk} \bar{T}_{ij(i)k(i,j).} \bar{T}'_{ij(i)k(i,j).}$$

$$\underline{S}_e(T) = \sum_{ijkl} T_{ij(i)k(i,j)l} T'_{ij(i)k(i,j)l} - \sum_{ijk} n_{ijk} \bar{T}_{ij(i)k(i,j).} \bar{T}'_{ij(i)k(i,j}$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{ijkl} T_{ij(i)k(i,j)l} T'_{ij(i)k(i,j)l}$$

mit den üblichen Invarianzeigenschaften, die ebenfalls für $\bar{Y}_{ij(i)..}$ für alle (i,j(i)) gelten. Nur die $\prod_{ij}(n_{ij.}!/\prod_k n_{ijk}!)$ Permutationen zwischen verschiedenen Stufen von Faktor C können zu unterschiedlichen Werten der Teststatistik führen.

pCRH-rc(s): Man hat Beobachtungen $Y_{ijk(i,j)l}$ mit $\hat{Q}^*_n = \times\times \hat{S}_{n_{ij.}}$ - invarianten Fehlern. In Analogie zu obigem Ergebnis hat man

(4.40)
$$T_{ijk(i,j)l} = Y_{ijk(i,j)l} - \bar{Y}_{ij..}$$

und dieselben Statistiken wie in (4.39), wenn dort stets j(i) durch j und c(i) durch c ersetzt wird.

pCRH-r(c)s: Wie im univariaten Fall ist ein Test auf Effekte des genesteten Faktors nicht möglich.

(ix) Lateinische Quadrate (pLS-k)

Die Auswahl eines lateinischen Quadrates wurde bereits in Abschnitt 3.1.5 (ix) erläutert. Die p-dimensionalen Zufallsvektoren sind $Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{k(i,j)} + \varepsilon_{ij}$ -i,j,k=1,..,l - mit $\hat{Q}^*_{l^2} = \hat{S}_{l^2}$ - invariantem Fehler. Die Transformationsmatrix zur Eliminierung der Störfaktoren A und B ist

$\hat{K} = (I_1 - 1/1J_1) \otimes (I_1 - 1/1J_1) \otimes I_p$, sodaß sich

(4.41) $T_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$ ergibt.

Bei Gültigkeit der Nullhypothese 'keine Effekte des Faktors C' ist $T = (T'_{11},\ldots,T'_{ll})'$ $\hat{Q}_{1^2} = \hat{S}_1 \otimes \hat{S}_1$-invariant, d.h. gegenüber allen $(l!)^2$ Vertauschungen ganzer Zeilen und Spalten von T_{ij}, wenn diese wie die ursprünglichen Y_{ij} im Versuchsplan angeordnet sind. In Analogie zum univariaten Fall wählt man

(4.42) $\underline{S}_h(T) = \sum_{k=1}^{l} \bar{T}_k \bar{T}'_k$ mit $\bar{T}_k = 1/1 \sum_{k(i,j)=k} T_{ij}$, denn $\bar{T}_{..} = 0$

und weiterhin (vgl. Kirk, 1968, S. 157 mit n=1 Replikationen je Zelle des Lateinischen Quadrates) mit $\bar{T}_{i.} = \bar{T}_{.j} = 0$ sowie $\sum_{ij} T_{ij} = \sum_{k=1}^{l} \sum_{k(i,j)=k} T_{ij} = \sum_{k=1}^{l} T_k$:

$$\underline{S}_e(T) = \sum_{i=1}^{l} \sum_{j=1}^{l} T_{ij} T'_{ij} - 1 \sum_{k=1}^{l} \bar{T}_k \bar{T}'_k$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{i=1}^{l} \sum_{j=1}^{l} T_{ij} T'_{ij} = \sum_{k=1}^{l} \sum_{k(i,j)=k} T_{ij} T'_{ij}$$

$\underline{S}_h + \underline{S}_e$ und der erste Term von $\underline{S}_e$ sind $\hat{Q}_{1^2}$-invariant wie auch $\bar{Y}_{i.}, \bar{Y}_{.j}, \bar{Y}_{..}$ für alle i,j in (4.41).

(x) Zweifaktorieller vollständiger Zufallsplan (pCRF-rc)

Ein Wechselwirkungstest existiert analog zum univariaten Testproblem nicht. Die Hypothese: 'Faktor B (bzw. A) hat überhaupt keinen Einfluß', d.h. auf keiner Stufe des Faktors A (bzw. B) (vgl. Abschnitt 3.1.5 (x)) ist wiederum testbar. Die Beobachtungen $Y_{ijk} = \mu + \beta_{ij} + \varepsilon_{ijk}$ mit i=1,..,r Stufen von Faktor A, j=1,..,c Stufen von Faktor B, k=1,..,n_{ij} > 1 Replikationen; $n = \sum\sum n_{ij}$ haben $\tilde{\Omega}_n = \tilde{S}_n$-invariante

Fehler. Die mit $\hat{\underline{K}}=\bigoplus_{i=1}^{r}((I_{n_i.}-1/n_i.J_{n_i.})\otimes I_p)$ gewonnenen transformierten Beobachtungen:

(4.43) $T_{ijk} = Y_{ijk} - \bar{Y}_{i..}$

sind $\hat{Q}_n=\hat{S}_r\sim\hat{S}_{n_i.}$ invariant, falls die Nullhypothese

H: $\beta_{i1}=..=\beta_{ic}$ für jedes i=1,..,r gilt.

Zur Konstruktion einer Teststatistik benötigt man:

(4.44) $\underline{S}_h(T) = \sum\sum_{ij} n_{ij}\bar{T}_{ij.}\bar{T}'_{ij.}$

$$\underline{S}_e(T) = \sum\sum\sum_{ijk} T_{ijk}T'_{ijk} - \sum\sum_{ij} n_{ij}\bar{T}_{ij.}\bar{T}'_{ij.} \qquad \text{sowie}$$

$(\underline{S}_h+\underline{S}_e)(T) = \sum\sum\sum_{ijk} T_{ijk}T'_{ijk}$ mit den üblichen Permutations-Invarianzeigenschaften von $\underline{S}_h+\underline{S}_e$ sowie dem ersten Summanden in $\underline{S}_e$. Außerdem ist $\bar{Y}_{i..}$ invariant für alle zulässigen Permutationen.

Dieselben Statistiken mit $T_{ijk} = Y_{ijk} - \bar{Y}_{.j.}$ verwendet man analog zum Testen von

H: $\alpha_{1j}=...=\alpha_{rj}$ für alle j=1,..,c.

In den folgenden Abschnitten werden vornehmlich Tests auf Wechselwirkungen zwischen zwei oder drei Faktoren von Interesse sein. Der Split-Plot Plan wurde bereits in Abschnitt 4.4 behandelt.

(xi) Zweifaktorieller vollständiger Blockplan (pRBF-rc)

Die p-dimensionalen Zufallsvektoren haben die Struktur

$Y_{hij}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\pi_h+\varepsilon_{hij}$ mit h=1,..,n Blöcken, i=1,..,r Stufen A, j=1,..,c Stufen Faktor B, mit $\hat{Q}_N^*=\hat{S}_{rc}\sim\hat{S}_{n.}$ - invarianten Fehlern, N=nrc.

Zum Testen der Wechselwirkungshypothese

$$H_0: (\alpha\beta)_{11} = \ldots = (\alpha\beta)_{rc} = 0$$

werden die Y_{hij} mit $\tilde{K} = I_n \otimes (I_r - 1/rJ_r) \otimes (I_c - 1/cJ_c) \otimes I_n$ transformiert zu:

$$(4.45) \quad T_{hij} = Y_{hij} - \bar{Y}_{hi.} - \bar{Y}_{h.j} + \bar{Y}_{h..} \quad .$$

Bei Gültigkeit der Nullhypothese sind diese $\tilde{Q}_N = (\tilde{S}_r \otimes \tilde{S}_c) \sim \tilde{S}_n$ - invariant, d.h. invariant gegen Vertauschungen ganzer Zeilen sowie Spalten je Block unabhängig von Block zu Block und Permutationen der ganzen Blöcke. Zur Konstruktion einer Teststatistik verwendet man:

$$(4.46) \quad \underline{S}_h(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n \bar{T}_{.ij} \bar{T}'_{.ij} \qquad (\bar{T}_{.i.} = \bar{T}_{..j} = \bar{T}_{...} = 0)$$

$$\underline{S}_e(T) = \sum_{h=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{c} T_{hij} T'_{hij} - \sum_{i=1}^{r} \sum_{j=1}^{c} n \bar{T}_{.ij} \bar{T}'_{.ij} \text{ , da auch } \bar{T}_{h..} = 0$$

$$(\underline{S}_h + \underline{S}_e)(T) = \sum_{h=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{c} T_{hij} T'_{hij} \quad .$$

$\underline{S}_h + \underline{S}_e$ sowie der erste Summand in $\underline{S}_e$ sind invariant unter den zulässigen Permutationen.

(xii) Dreifaktorielle Versuchspläne, Test auf dreifache Wechselwirkung

Wie im univariaten Fall gibt es keinen Test auf dreifache Interaktion im dreifaktoriellen Zufallsplan (pCRF-rcs) sowie im Split-Plot-Plot Plan (pSPF-rc.s).

Die Teststatistiken für den multivariaten dreifaktoriellen vollständigen Blockplan (pRBF-rcs) wie für den multivariaten Split-Split-Plot Plan (pSPF-r.cs) sind aus den entsprechenden univariaten Plänen wieder direkt herleitbar.

RBE-rcs: Die Modellgleichung für p-variate Beobachtungen entspricht (3.74):

$$Y_{hijk}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk}+\pi_h+\varepsilon_{hijk}$$

mit h=1,..,n Blöcken; i=1,..,r Stufen Faktor A; j=1,.., c Stufen Faktor B; k=1,..,s Stufen Faktor C; N=nrcs und mit $\tilde{Q}_N^*=\tilde{S}_{rcs}\sim\tilde{S}_n$-invariantem Fehler.

Zum Testen von H: $(\alpha\beta\gamma)_{111}=\ldots=(\alpha\beta\gamma)_{rcs}=0$ wird zuerst der Beobachtungsvektor Y mit $\tilde{K}=I_n\otimes(E_r\otimes E_c\otimes E_s)\otimes I_p$ mit $E_\mu=(I_\mu-1/\mu J_\mu)$ transformiert:

(4.47) $$T_{hijk}=Y_{hijk}-\bar{Y}_{hij.}-\bar{Y}_{hi.k}-\bar{Y}_{h.jk}+\bar{Y}_{hi..}+\bar{Y}_{h.j.}+\bar{Y}_{h..k}-\bar{Y}_{h...}$$

Bei Gültigkeit der Nullhypothese ist die gemeinsame Verteilung der T_{hijk} $\tilde{Q}_N=(\tilde{S}_r\otimes\tilde{S}_c\otimes\tilde{S}_s)\sim\tilde{S}_n$ - invariant.
Als Bestandteile der Teststatistiken hat man:

(4.48) $$\underline{S}_h(T)=\sum_{ijk}\sum n\bar{T}_{.ijk}\bar{T}'_{.ijk}\quad,\ \text{da}\ \bar{T}_{.ij.}=\bar{T}_{.i.k}=\bar{T}_{..jk}=\bar{T}_{.i..}=\bar{T}_{..j.}=\bar{T}_{...k}=\bar{T}_{....}=0\ ;$$

$$\underline{S}_e(T)=\sum_{hijk}\sum T_{hijk}T'_{hijk}-\sum_{ijk}\sum n\bar{T}_{.ijk}\bar{T}'_{.ijk}$$

(da auch $\bar{T}_{....}=0$)

$$(\underline{S}_h+\underline{S}_e)(T)=\sum_{hijk}\sum T_{hijk}T'_{hijk}$$

mit den üblichen Invarianzeigenschaften von $\underline{S}_h+\underline{S}_e$ und dem ersten Summanden von $\underline{S}_e$.

pSPF-rcs: Mit der Modellgleichung analog zu (3.79) und Kirk (1968, S. 299) in p abhängigen Variablen sowie $\tilde{Q}_N^*=\tilde{S}_{cs}\sim\tilde{S}_n$ -invariantem Fehler werden zum Testen von H: $(\alpha\beta\gamma)_{ijk}=0$ für alle i,j,k die Zufallsvektoren wie folgt mit $K=E_n\otimes E_c\otimes E_s\otimes I_p$ transformiert:

(4.49) $$T_{ih(i)jk}=Y_{ih(i)jk}-\bar{Y}_{ih(i)j.}-\bar{Y}_{ih(i).k}+\bar{Y}_{ih(i)..}-\bar{Y}_{..jk}+\bar{Y}_{..j.}+\bar{Y}_{...k}-\bar{Y}_{....}\quad.$$

Die gemeinsame Verteilung der T_{ihjk} ist bei Gültigkeit obiger Hypothese $\tilde{Q}_N=\tilde{S}_{n_\cdot}\otimes(\tilde{S}_c\otimes\tilde{S}_s)$ - invariant (vgl. 3.1.5 (xiii)).

Zur Konstruktion einer Teststatistik benötigt man wieder:

(4.50) $\underline{S}_h(T) = \sum\sum\sum_{ijk} n_i \bar{T}_{i.jk}\bar{T}'_{i.jk}$,da $\bar{T}_{i.j.}=\bar{T}_{i..k}=\bar{T}_{i...}=\bar{T}_{..j.}=\bar{T}_{...k}=\bar{T}_{....}=0$;

$$\underline{S}_e(T) = \sum\sum\sum\sum_{ihjk} T_{ih(i)jk}T'_{ih(i)jk} - \sum\sum\sum_{ijk} n_i \bar{T}_{i.jk}\bar{T}'_{i.jk} \text{ , da } \bar{T}_{ih(i)j.}=\bar{T}_{ih(i).k}=\bar{T}_{ih(i)..}=0$$

$$(\underline{S}_h+\underline{S}_e)(T) = \sum\sum\sum\sum_{ihjk} T_{ih(i)jk}T'_{ih(i)jk} \quad .$$

Nur $\underline{S}_h$ und der zweite Summand von $\underline{S}_e$ sind nicht invariant unter den zulässigen Vektor-Permutationen aus $\tilde{\Omega}_N$.

Wiederum kann man wie beim p-variaten Split-Plot Plan für alle drei bei univariaten Beobachtungen unterschiedenen Fälle (a) - (c) (vgl. Abschnitt 3.1.5 (xiii)) im multivariaten Plan eine quadratische Form als Teststatistik konstruieren. In offensichtlicher Verallgemeinerung der univariaten Resultate findet man mit $T_N=(\bar{T}'_{1.11},\ldots,\bar{T}'_{r.cs})'$ und $\tilde{P}_N$ die p-dimensionale Verallgemeinerung von P_N:

Fall (a): Die Zuteilung zu den Faktorstufenkombinationen aller drei Faktoren kann randomisiert erfolgen:

(4.51) $$Q(T_N)=\sum_{l=1}^{p}\sum_{l'=1}^{p}\sigma^{*ll'}_{N,t}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n_i\bar{T}_{i.jkl}\bar{T}_{i.jkl'} = (n_\cdot-1)(c-1)(s-1)V^{(s)}$$

mit $\mathrm{Var}_{\tilde{P}_N}(T_{ihjk}) = (n_\cdot-1)(c-1)(s-1)\,n_\cdot cs\Sigma_{N,t}$ und $\sigma^{*ll'}_{N,t}$ Elemente der (verallgemeinerten) Inversen von $\Sigma_{N,t}=((\sigma^{ll'}_{N,t}))_{p\times p}$ mit

$$\sigma^{ll'}_{N,t} = 1/(n_\cdot-1)(c-1)(s-1)\sum\sum\sum\sum\sum_{ihjk} t_{ihjkl}\cdot t_{ihjkl'} \quad .$$

$V^{(s)}$ ist stets das aus $(\underline{S}_h+\underline{S}_e)^{-1}(T)$ und $\underline{S}_h(T)$ gebildete Pillai'sche Spurkriterium.

Fall (b): Nur die Zuteilung zu den Stufen des Plot-Faktors kann randomisiert erfolgen:

$$(4.52)\quad Q(T_N) = \sum_{i=1}^{r} n_i \sum_{j=1}^{c} \sum_{j'=1}^{c} \sum_{k=1}^{s} \sum_{k'=1}^{s} \sum_{l=1}^{p} \sum_{l'=1}^{p} \bar{T}_{i.jkl}\, \sigma^{*ll'}_{jj'kk'}\, \bar{T}_{i.j'k'l'} \quad .$$

Die $\sigma^{*ll'}_{jj'kk'}$ sind Elemente der (verallgemeinerten) Inversen von $\hat{\Sigma}_{n.} = ((\Sigma_{jj'kk'}))_{pcs \times pcs}$ mit $\Sigma_{jj'kk'} = ((\sigma^{ll'}_{jj'kk'}))_{p \times p}$ und $\sigma^{ll'}_{jj'kk'} = 1/(n_. - 1) \sum_{ih} t_{ihjkl} \cdot t_{ihj'k'l'}$

$l,l' = 1,..,p;\ j,j' = 1,..,c;$
$k,k' = 1,..,s.$

Fall (c): Die Zuteilung zu den Plot-Faktorstufen sowie den Stufen des ersten Split-Faktors sei randomisiert möglich:

$$(4.53)\quad Q(T_N) = \sum_{i=1}^{r} n_i \sum_{j=1}^{c} \sum_{k=1}^{s} \sum_{k'=1}^{s} \sum_{l=1}^{p} \sum_{l'=1}^{p} \bar{T}_{i.jkl}\, \sigma^{*ll'}_{kk'}\, \bar{T}_{i.jk'l'} \quad .$$

Die $\sigma^{*ll'}_{kk'}$ sind Elemente der (verallgemeinerten) Inversen von $\hat{\Sigma}_{n.c} = ((\Sigma_{kk'}))_{ps \times ps}$ mit $\Sigma_{kk'} = ((\sigma^{ll'}_{kk'}))_{p \times p}$ und $\sigma^{ll'}_{kk'} = 1/(n_. - 1)(c-1) \sum_{ihj} t_{ihjkl} \cdot t_{ihjk'l'}$.

4.6. Multivariate Permutationstests zur Analyse von Profilen

Der MANOVA-Ansatz zur Analyse des r-Stichprobenproblems bei wiederholten Messungen je Beobachtungseinheit wird oft als Profilanalyse bezeichnet (Greenhouse & Geisser, 1959; Morrison, 1976). Üblicherweise handelt es sich um wiederholte Beobachtungen zu p verschiedenen festen Zeitpunkten bezüglich derselben abhängigen Variablen oder um die erzielten Punktwerte in p Untertests einer psychologischen Testbatterie (Cole & Grizzle, 1966; Finn, 1969; Bock, 1975; Monlezun et al., 1984). Weiterhin wird angenommen, daß die wiederholten Messungen 'vergleichbar' (commensurate, vgl.

z.B. Morrison, 1976, S. 141) sein sollen, d.h. es soll sich um Messungen auf derselben Skala oder Skalen mit gleicher Lage und gleichen Einheiten handeln.

MANOVA-Verfahren sind immer dann erforderlich, wenn sog. 'mixed-model assumptions' der ANOVA nicht gelten (vgl. z.B. Kirk, 1968 Kap. 8.4.; Davidson, 1972; Timm, 1980, S. 54, Koch et al., 1980; Rogan, Keselman & Mendoza, 1979). Als Alternative zu MANOVA-Verfahren sind auch univariate approximative Tests vorgeschlagen worden (Geisser & Greenhouse, 1958; Huynh & Feldt, 1970; Huynh, 1978).

(i) Von besonderem Interesse ist der Test auf Parallelität der Profile unter r experimentellen Bedingungen. Mit Hilfe des allgemeinen linearen Modells läßt sich die Hypothese der Profilparallelität leicht formulieren.

Sei $Y = \begin{pmatrix} Y'_{11} \\ Y'_{rn_r} \end{pmatrix}$ die $n_. \times p$-Matrix der p-dimensionalen zeilenweise angeordneten Beobachtungsvektoren mit unabhängigen und im parametrischen Fall $NV_p(\mu_i, \Sigma)$ - verteilten Y'_{ih} , $i=1,..,r$; $h=1,..,n_i$; $n_. = \sum_{i=1}^{r} n_i$.

Für den Erwartungswert von Y gelte

$$E(Y) = X \cdot B$$

mit der $n_. \times r$-dimensionalen Design-Matrix $X = \bigoplus_{i=1}^{r} 1_{n_i}$ und der $r \times p$-dimensionalen Parameter-Matrix $B = \begin{pmatrix} \mu'_1 \\ \mu'_r \end{pmatrix}$ und $Var(Y) = I_{n_.} \otimes \Sigma$ im NV_p-Fall.

Allgemeine lineare Hypothesen lassen sich als

$$H : CBA = \Gamma$$

darstellen (d.h. rg(C) unabhängige, schätzbare parametrische Vektor-Funktionen $\psi_i = c'_i B$ haben den festen Wert Γ).

Für den Test auf Profilparallelität mit $\Gamma = 0$ haben die beiden $(r-1) \times r$ bzw. $p \times (p-1)$-dimensionalen (Kontrast-) Matrizen C und A i.a. folgende Gestalt (vgl. Timm, 1980, S. 51;

Morrison, 1976, S. 188; Kshirsagar, 1972):

(4.54) $C = (I_{r-1} | -1_{r-1})$ $\qquad A = I_{p-1} \otimes (1,-1)'$.

Prinzipiell ist auch jede andere Kontrast-Matrix C bzw. A mit linear unabhängigen Zeilen bzw. Spalten wählbar. Eine (k-1)xk bzw. kx(k-1) Matrix D heißt Zeilen/Spalten-Kontrastmatrix, falls $D \cdot 1_k = 0$ bzw. $1_k \cdot D = 0$.
Falls A wie in (4.54) gewählt wird, werden jeweils die Differenzen benachbarter Profilstudien gebildet. Der Vergleich einer (ausgewählten) Profilstufe wie z.B. erster/letzter Zeitpunkt als Kontrolle, Vergleichs-oder Ankertest bei einer Testbatterie mit jeder der übrigen Stufen, d.h. z.B. $A = (I_{p-1} | -1_{p-1})'$ ist oft inhaltlich ebenfalls sinnvoll. Zusätzlich muß natürlich stets sichergestellt sein, daß Differenzen von Profilwerten sinnvoll definiert sind (vgl. z.B. Koch et al., 1980, S. 256).

In der MANOVA sind die Teststatistiken, wie zuvor dargestellt, Funktionen von:

(4.55) $\underline{S}_h(Y) = (C(X'X)^{-}X'YA-\Gamma)'(C(X'X)^{-}C')^{-1}(CX'X)^{-}X'YA-\Gamma)$

$\underline{S}_e(Y) = A'Y'(I-X(X'X)^{-}X')YA$ (vgl. z.B. Timm, 1980)

Setzt man $Z = YA$, sieht man sofort, daß gilt:

$$\underline{S}_h(Z) = A'\underline{S}_h(Y)A \quad \text{und} \quad \underline{S}_e(Z) = A'\underline{S}_e(Y)A \quad .$$

$\underline{S}_h(Y)$ und $\underline{S}_e(Y)$ sind aber die Hypothesen und Fehlermatrizen für das p-dimensionale unverbundene r-Stichprobenproblem (vgl. Morrison, 1976, S. 188). Der Test auf Profilparallelität ist also identisch mit einem Test auf Gleichheit der r Bedingungen in den intraindividuell transformierten Beobachtungen (Profilkontrasten).

Falls A die Eigenschaft $A'A = I_{p-1}$ hat und Σ die Bedingung der r-Stichproben-Sphärizität (Hager & Westermann, 1983a, S. 140) erfüllt, gilt für die S.d.q.A. in Zähler und Nenner der F-Statistik beim Wechselwirkungstest im Split-Plot Design:

Zähler: $sp(\underline{S}_h(Z))$ mit $df=rg(C)rg(A)=(r-1)(p-1)$ und

Nenner: $sp(\underline{S}_e(Z))$ mit $df=(n_.-r)(p-1)$ (Timm, 1980,S.54)

(ii) Sind die unabhängigen Y_{ij} nicht normalverteilt, sondern nach (stetigem) $F_i(x) = G(x-\mu_i)$, hat Koch (1969) einen Rangtest auf Profilparallelität vorgeschlagen.

Für jedes Paar (j,j') - j,j'=1,..,p-von Profilelementen werden die $n_.$ Differenzen $Y_{ihj} - Y_{ihj'}$, i=1,..,r; h=1,..,n_i, gebildet und jeweils durch ihren Totalrang $R_{ih}^{jj'}$ bzgl. aller $n_.$ Differenzen ersetzt. Anschließend wird je Profilstufe j=1,..,p die Rangsumme $R_{ih}^{j} = \sum_{j'=1}^{p} R_{ih}^{jj'}$ bestimmt. Sei $R_{ih}' = (R_{ih}^1,..,R_{ih}^p)$.

Als Teststatistik wählt Koch (1969, Formel (34)) eine quadratische Form in den Vektoren $\bar{R}_{i.}$ mit $\bar{R}_{i.}^{j} = 1/n_i \sum_{h=1}^{n_i} R_{ih}^{j}$,j=1,..,p

(4.56) $L_{n_.} = (n_.-1)/n_. \cdot \sum_{i=1}^{r} n_i (C\bar{R}_{i.})'(CV_{n_.}^{*}C')^{-1}C\bar{R}_{i.}$

und der Kontrastmatrix C aus (4.54) sowie der Varianz-Kovarianz-Matrix in den $\bar{R}_{i.}$:

$V_{n_.} = ((v_{jj'}))_{p\times p}$ mit $v_{jj'} = 1/n_. \sum_{ih}(\bar{R}_{ih}^{j} - p(n_.+1)/2)(\bar{R}_{ih}^{j'} - p(n_.+1)/2)$

$L_{n_.}$ hat asymptotisch ($n_. \to \infty$) eine χ^2-Verteilung mit $df=(r-1)(p-1)$. Es kann aber auch die exakte (bedingte) Permutationsverteilung unter dem Permutationsmaß $P_{n_.}$ (vgl. Fall (b) in Abschnitt 3.1.5) mit $n_.!/\Pi n_i!$ Vertauschungen der $n_.$ p-dimensionalen Rangsummen-Vektoren R_{ih} benutzt werden. Damit ist $L_{n_.}$ eine nichtparametrische Alternative zur MANOVA-Teststatistik

$sp(\underline{S}_h \cdot \underline{S}_e^{-1})$ mit $\underline{S}_h$ und $\underline{S}_e$ aus (4.55) .

Bhapkar & Patterson (1977, 1978) geben eine Klasse von nichtparametrischen Tests basierend auf U-Statistiken für dasselbe Testproblem an. Sie weisen darauf hin, daß zum Test der obigen Hypothese für Verteilungen $F_i(x)=G(x-\mu_i)$, $x,\mu_i \in R^p$ alle eindimensionalen Randverteilungen $G^l(x)$, $l=1,..,p$, identisch sein müssen. Diese Forderung hat große Ähnlichkeit mit der zuvor erwähnten 'Vergleichbarkeit' (commensurability) der Profilvariablen.

(iii) Betrachtet man die Formulierung der Profilparallelitätshypothesen für univariate Invarianz-Permutationstests in (3.63), handelt es sich links vom Gleichheitszeichen gerade um die Anwendung der Kontrastmatrix A aus (4.54) auf die p-dimensionalen individuellen Profile $\mu_{ih}=(\mu_{ih1},..,\mu_{ihp})'$ mit p=c. Die Vertauschbarkeitsannahme für die Fehlerprofile war dort dieselbe wie für das p-dimensionale unverbundene Mehrstichprobenproblem in Abschnitt 4.5 (iii).

Bei Bildung der Kontraste je individuellem Profil werden mögliche Niveauunterschiede der Profile innerhalb jeder Faktorstufe i=1,..,r eliminiert. Die Nullhypothese der Profilparallelität über alle Faktorstufen i=1,..,r hinweg ist dann identisch mit der Nullhypothese im (p-1)MCR-r für die (p-1)-dimensionalen Kontraste in den Profilstufen. Die Permutationsinvarianzeigenschaften der gemeinsamen Verteilung F(x) der Fehlerprofile aus (3.62), d.h. $F(x)=F(\pi x)$ für alle $\pi \in \tilde{S}_{n.} = S_{n.} \otimes \{1_p\}$, wird durch die Bildung von Kontrasten innerhalb der Fehlerprofile nicht verändert, d.h. $(I_{n.} \otimes A')\varepsilon$ hat eine Verteilungsfunktion F^A mit der Eigenschaft

$$F^A(x) = F^A(\pi x) \text{ für alle } \pi \in S_{n.} \otimes \{1_{p-1}\} \quad .$$

Mit $Z=(I_{n_.} \otimes A')\cdot Y$ betrachtet man wie in (4.29):

$$(4.57)\quad \underline{S}_h(Z) = \sum_{i=1}^{r} n_i(\bar{\bar{z}}_{i.}-\bar{\bar{z}}_{..})(\bar{\bar{z}}_{i.}-\bar{\bar{z}}_{..})'$$
$$= \sum_{i} n_{.}(A'\bar{\bar{y}}_{i.}-A'\bar{\bar{y}}_{..})(A'\bar{\bar{y}}_{i.}-A'\bar{\bar{y}}_{..})'$$
$$= A'\underline{S}_h(Y)A \qquad \text{und analog}$$
$$\underline{S}_e(Z) = A'\underline{S}_e(Y)A$$
$$(\underline{S}_h+\underline{S}_e)(Z) = A'((\underline{S}_h+\underline{S}_e)(Y))A \quad .$$

Wiederum ist $(\underline{S}_h+\underline{S}_e)(Z)$ invariant unter den bei Gültigkeit der Profilparallelitätshypothese zulässigen Permutationen aller $Z_{ih} = (Y_{ih1}-Y_{ih2}, Y_{ih2}-Y_{ih3}, \ldots, Y_{ih(p-1)}-Y_{ihp})'$. Von diesen können wiederum nur $n_{.}!/\prod_i n_i!$ $(p-1)$-dimensionale Vektorpermutationen zu verschiedenen Werten in einer aus $\underline{S}_h(Z)$ und $\underline{S}_e(Z)$ gebildeten Teststatistik führen.

Anstelle von $A=I_{p-1}\otimes(1,-1)'$ hätten auch andere Matrizen zur Bildung von Kontrasten der Profilelemente benutzt werden können. Mit

$$(4.58)\quad A^{O} = (I_{p-1}|-1_{p-1})'$$

etwa erhält man alle Kontraste mit dem letzten Profilelement (vgl. z.B. Kshirsagar, 1972, S. 173). Mit der sog. Helmert'schen Orthogonal-Matrix $\tilde{A}$ wird sukzessive der Mittelwert aus den vorhergehenden Profilelementen mit dem jeweils letzten Profilelement kontrastiert:

$$\tilde{A}_{p\times(p-1)} = \begin{pmatrix} 1/d_1 & 1/d_2 & & 1/d_{p-1} \\ -1/d_1 & 1/d_2 & \ldots & 1/d_{p-1} \\ 0 & -2/d_2 & \ldots & 1/d_{p-1} \\ & & & \\ 0 & 0 & \ldots & -(p-1)/d_{p-1} \end{pmatrix} \quad \text{mit } d_i=(i(i+1))^{1/2},\ i=1,..,p-1$$

Die MANOVA-Teststatistiken in (4.18), gebildet mit $\underline{S}_h$ und $\underline{S}_e$ aus (4.55) sind jedoch invariant gegen die spezielle Wahl einer Kontrastmatrix A. Zwei mögliche $p\times(p-1)$ Kontrastmatrizen A_α und A_β sind stets durch eine $(p-1)\times(p-1)$ Matrix D von vollem Rang ineinander überführbar (vgl. z.B.

Khsirsagar, 1972, S. 153-155):

$A_\alpha = A_\beta D$ und es gilt mit den Bezeichnungen aus (4.18):

$$V^{(s)}(Y,A_\alpha) = V^{(s)}(Y,A_\beta) \; ; \; V(Y,A_\alpha) = V(Y,A_\beta)$$

$$\Lambda(Y,A_\alpha) = \Lambda(Y,A_\beta) \; ; \; U(Y,A_\alpha) = U(Y,A_\beta) \; .$$

Eine weitere Vereinfachung von $\underline{S}_h(Y_A)$ ist jedoch nicht möglich, da A nicht quadratisch ist. Ein Permutationstest auf Niveauunterschiede zwischen den Profilstufen in Analogie zum parametrischen Fall (Timm, 1980, S. 51) ist nicht möglich, da nur Permutationen ganzer Profile zulässig sind. Diese lassen jedoch jede geeignete Teststatistik invariant. Nur bei einem Modellwechsel, in dem auch S_p-Invarianz der Verteilungsfunktion der Fehlerprofile gegeben ist, könnte auf Lageunterschiede zwischen den Profilstufen separat für jede Plotstufe oder bei Parallelität der Profile über alle $n_.$ Beobachtungseinheiten hinweg ein Permutationstest wie im RB-p Plan benutzt werden.

(iv) Auch zum Test auf verallgemeinerte Profilparallelität im Split-Split-Plot Design (vgl. Kap. 3.1.5 (xiii)) kann man einen multivariaten Permutationstest angeben. Dabei wird vorausgesetzt, daß alle cs Profilelemente vergleichbare Messungen darstellen. Man hat in i=1,..,r Gruppen je n_i verallgemeinerte cs-dimensionale Profile μ_{ihjk}, h=1,..,n_i. Die verallgemeinerte Profilparallelitätshypothese (vgl. (3.83)):

$$\mu_{ihjk} - \mu_{ihj,k+1} - \mu_{ihj+1,k} + \mu_{ihj+1,k+1} = \Delta_{jk}$$ unabhängig

von der Faktorstufe (Gruppe) i des Plotfaktors, läßt sich auffassen als Anwendung einer verallgemeinerten csx(c-1)(s-1) dimensionalen Kontrastmatrix A auf die individuellen cs-dimensionalen Profile μ_{ih}. Sei

(4.60) $B_{csxc-1} = I_{c-1} \otimes (1,-1)' \otimes 1_s = (b_1,\ldots,b_{c-1})$ eine Kontrastmatrix bzgl. des Index j (erster Splitfaktor),

$D_{csxs-1} = 1_c \otimes I_{s-1} \otimes (1,-1)' = (d_1,\ldots,d_{s-1})$ eine Kontrastmatrix bzgl. Index k (zweiter Splitfaktor).

Dann ist die gesuchte verallgemeinerte Kontrastmatrix

$A:=B*D := (b_1*d_1,..,b_1*d_{s-1},b_2*d_1,..,b_{c-1}*d_1,..,b_{c-1}*d_{s-1})$

mit der Dimension csx(c-1)(s-1) sowie b_i*d_j dem Hadamard-Produkt der beiden Vektoren b_i und d_j:

$b_i*d_j = (b_{i1}d_{j1},..,b_{ics}d_{jcs})$

(vgl. Timm, 1980, S. 59).

Bei Bildung der Kontraste je individuellem Profil verschwinden möglicherweise bestehende Niveauunterschiede der Profile innerhalb jeder Faktorstufe i=1,..,r. Die Nullhypothese der verallgemeinerten Profilparallelitat ist damit identisch mit der Nullhypothese im (c-1)(s-1)-dimensionalen MCR-r für die (c-1)(s-1)-dimensionalen verallgemeinerten Profilkontraste. Die Permutations-Invarianzeigenschaften der Verteilungsfunktion der Fehlerprofile ε mit $F(x) = F(\pi x)$ für alle $\pi \in S_{n.} \wr \{1_{cs}\}$ wird durch die Transformation $(I_{n.} \otimes A')\varepsilon$ nicht wesentlich verändert:

$F^A(x) = F^A(\pi x)$ für alle $\pi \in S_{n.} \otimes \{1_{(c-1)(s-1)}\}$, $x \in R^{n.(c-1)(s-1)}$.

Mit $Z = (I_{n.} \otimes A')Y$ werden analog zu (4.57) $\underline{S}_h(Z)$ und $\underline{S}_e(Z)$ gebildet, und man verfährt zur Durchfuhrung des Permutationstests wie zuvor beschrieben.

Permutationstests auf Wechselwirkung zwischen den beiden 'Profil-Faktoren' bzw. auf deren Haupteffekte wie in Timm (1980, S. 59f) sind wegen der Permutationsinvarianz der MANOVA-Teststatistiken wiederum nicht möglich. In Koch (1970, Abschnitt 3.7) ist ein Rangtest für dieses Problem

angegeben.

Falls verallgemeinerte Profilparallelität angenommen werden kann, ist ein Test auf Profilparallelität bzgl. jedes der beiden Profilfaktoren für sich genommen möglich, indem man als csx(c-1) bzw. csx(s-1)-dimensionale Kontrastmatrix B oder D aus (4.60) benutzt. Anwendung von B bzw. D auf die individuellen Profile μ_{ihjk} ergibt nämlich:

$\mu_{ihj.}-\mu_{ih,j+1.}$ bzw. $\mu_{ih.k}-\mu_{ih.,k+1}$.

Anstelle von B und D in (4.60) hätten auch in Analogie zu den Überlegungen vorher andere Kontrastmatrizen zur Bildung von A= B*D benutzt werden können wie:

(4.61) $B^{\circ} = (I_{c-1}|-1_{c-1})'\otimes 1_s$ und $D^{\circ} = 1_c\otimes(I_{s-1}|-1_{s-1})'$

oder zwei Helmert-Matrizen $\tilde{B} = \tilde{A}_{cx(c-1)}\otimes 1_s$ und $\tilde{D} = 1_c\otimes\tilde{A}_{sx(s-1)}$ mit $\tilde{A}$ aus (4.59).

Im Fall $A^{\circ}=B^{\circ}*D^{\circ}$ haben die verallgemeinerten Kontraste folgendes Aussehen:

$\mu_{ihjk} - \mu_{ihjs} - \mu_{ihck} - \mu_{ihcs}$,j=1,..,c-1; k=1,..,s-1,

für alle i=1,..,r; h=1,.., n_i.

Die spezielle Wahl der Kontrastmatrizen verändert jedoch wiederum die Matrizen $\underline{S}_e$ und $\underline{S}_h$ nicht.

(v) Falls man sog. multivariate wiederholte Messungen (Timm, 1980, S. 68ff; Bock, 1975, Kap. 7.3.4) vorliegen hat, d.h. Beobachtungen bzgl. p verschiedenen Variablen über c Bedingungen (Zeitpunkte) hinweg für r Gruppen, kann man auch anstelle des Tests für Fall (b) in Abschnitt 4.4 eine Profilanalyse ausführen. Vergleichbar sind nur die Beobachtungen bzgl. jeder einzelnen der p Variablen über die c Bedingungen. Ordnet man o.B.d.A. die Beobachtungen je

Beobachtungseinheit zuerst nach Bedingungen (j), dann nach Variablen (l):

$Y_{ih}=(Y_{ih11},\dots,Y_{ihc1},\dots,Y_{ih1p},\dots,Y_{ihcp})'$, ist die Hypothese der Profilparallelität für jede der p Variablen über die r Gruppen hinweg:

(4.62) H: $\mu_{ihjl}-\mu_{ih,j+1,l}=\Delta_{jl}$ für alle i,h(i)

gegen

A: $\mu_{ihjl}-\mu_{ih,j+1,l}=\Delta_{ijl}$ für alle h(i) und

$\Delta_{ijl}\neq\Delta_{i'jl}$ für ein $i\neq i'$ und ein $l\in\{1,..,p\}$

Diese Hypothese läßt sich auch darstellen mit Hilfe der Kontrastmatrix

(4.63) $\overset{\circ}{A}_{pc\times p(c-1)}=I_p\otimes A_{c\times(c-1)}$ und A eine der Kontrastmatrizen aus (4.54), (4.56); sowie $\overset{\circ}{C}=C$ wie in (4.54).

Wie zuvor impliziert die $S_{n.}\otimes\{1_{cp}\}$-Invarianz der Verteilungsfunktion der Fehlerprofile die $S_{n.}\otimes\{1_{p(c-1)}\}$-Invarianz der Verteilungsfunktion von $(I_{n.}\otimes\overset{\circ}{A}')Y$.

Damit kann man zur Bildung einer Teststatistik für einen Permutationstest mit $Z=(I_{n.}\otimes\overset{\circ}{A}')Y$, wobei die Y_{ih} angeordnet sind wie oben, die bekannten Statistiken

$$\underline{S}_h(Z)=\overset{\circ}{A}'\underline{S}_h(Y)\overset{\circ}{A}\quad,\quad \underline{S}_e(Z)=\overset{\circ}{A}'\underline{S}_e(Y)\overset{\circ}{A}$$

benutzen. $\underline{S}_h$, $\underline{S}_e$ sind jetzt p(c-1)xp(c-1)-dimensionale Matrizen. $(\underline{S}_h+\underline{S}_e)(Z)$ und der Term $n_{.}\overset{\circ}{A}'\bar{Y}_{..}\bar{Y}_{..}'\overset{\circ}{A}$ sind invariant unter den n.! bei Gültigkeit von H aus (4.62) zulässigen Permutationen der p(c-1)-dimensionalen

$$Z_{ih}=(Y_{ih11}-Y_{ih12},Y_{ih12}-Y_{ih13},\dots,Y_{ihp,c-1}-Y_{ihpc})'\ ;$$

von diesen können wiederum nur $n_{.}!/\prod_i n_i!$ zu verschiedenen Teststatistikwerten führen.

Ein Permutationstest auf Unterschiede zwischen den c Faktor-Stufen bzgl. der p Variablen ist nicht möglich, da nur Permutationen ganzer pc-dimensionaler Profile zulässig sind. Permutationstests wie im p-dimensionalen RB-c sind

nur bei Modellwechsel, d.h. $\tilde{\tilde{S}}_c = \tilde{\tilde{S}}_c \otimes \{1_p\}$ -Invarianz der Verteilungsfunktion der Fehlerprofile möglich.

(vi) Den Ansatz aus (v) kann man auf den Fall des p-dimensionalen Split-Split-Plot Planes (pSPF-r.cs, vgl. Abschnitt 4.4 (xii)) verallgemeinern. In diesem Design hat man zwei Faktoren mit c bzw. s wiederholten Messungen bezüglich p nicht untereinander vergleichbarer Variablen. Die Beobachtungen über alle cs-Faktorstufenkombinationen seien aber je Variable vergleichbar. Für die Fehler-Verteilungsfunktion sei abschwächend nur $\tilde{Q}_N^* = S_{n.} \otimes \{1_{csp}\}$-Invarianz angenommen gegenüber der $\tilde{S}_{cs} \tilde{S}_{n.}$ -Invarianz in Abschnitt 4.4 (xii).

Ordnet man erneut die Beobachtungen je Beobachtungseinheit zuerst nach den Bedingungskombinationen (j,k), dann nach den Variablen (l):

$$Y_{ih} = (Y_{ih111},..,Y_{ihcs1},Y_{ih112},..,Y_{ih11p},..,Y_{ihcsp})' ,$$

ist die Hypothese der verallgemeinerten Profilparallelität für jede der p Variablen über die r Gruppen hinweg:

(4.64) H: $\mu_{ihjkl} - \mu_{ihj,k+1,l} - \mu_{ih,j+1,kl} + \mu_{ih,j+1,k+1,l} = \Delta_{jkl}$ unabhängig von den (Plot-)Stufen i=1,..,r für alle h=1,..,n_i; l=1,..,p.

Diese Hypothese läßt sich wiederum darstellen als Anwendung der verallgemeinerten Kontrastmatrix:

(4.65) $\tilde{\tilde{A}}_{pcs \times p(c-1)(s-1)} = I_p \otimes B^* D$ mit B,D aus (4.60)

auf die individuellen pcs-dimensionalen Profile μ_{ih}. Bei Bildung des Kontrastes je individuellem Profil verschwinden mögliche Niveauunterschiede der Profile innerhalb jeder Faktorstufe i=1,..,r. Die Nullhypothese der verallgemeinerten Profilparallelität ist damit identisch mit der Null-

hypothese im $p(c-1)(s-1)$-dimensionalen MCR-r für die $p(c-1)(s-1))$-dimensionalen verallgemeinerten Profilkontraste.

Die Verteilungsfunktion von $(I_{n_\bullet} \otimes \tilde{A}')\varepsilon$ besitzt unter H aus (4.64) $S_{n_\bullet} \otimes \{1_{p(c-1)(s-1)}\}$-Invarianz. Mit $Z=(I_{n_\bullet} \otimes \tilde{A}')Y$ werden wiederum analog zu (4.57) die Statistiken $\underline{S}_h(Z)$ und $\underline{S}_e(Z)$ gebildet. Der Permutationtest wird wie zuvor beschrieben durchgeführt. Jede der MANOVA-Teststatistiken ist wieder invariant gegen die spezielle Wahl der Kontraste B,D.

Permutationstests auf Wechselwirkung zwischen den beiden Profilfaktoren bzw. auf deren Haupteffekte sind wegen der $S_{n_\bullet} \otimes \{1_{p(c-1)(s-1)}\}$-Permutationsinvarianz der MANOVA-Teststatistiken nicht möglich, wohl aber unter Voraussetzung der verallgemeinerten Profilparallelität ein Test auf Profilparallelität bzgl. jedes einzelnen der beiden Profilfaktoren. Dazu benutzt man als $pcsx(c-1)$ bzw. $pcsx(s-1)$-dimensionale Kontrastmatrix $\tilde{B}=I_p \otimes B$, $\tilde{D}=I_p \otimes D$ aus (4.60). Wiederum können auch anstelle von B oder D, B° oder D°, $\tilde{B}$ oder $\tilde{D}$ aus (4.61) als Kontrastmatrizen verwendet werden.

4.7 Multivariate Permutations/Randomisierungstests mit normierten Distanzen zwischen Beobachtungen als Analyseeinheiten

4.7.1 Unverbundenes multivariates k-Stichprobenproblem

Die bisherigen Überlegungen zu multivariaten Permutationstests waren ausschließlich am linearen Modell der MANOVA ohne die restriktive Normalverteilungsannahme orientiert. Mielke et al. (1976), Mielke (1978) testen dagegen für den Fall von k Gruppen mit jeweils $n_i, i=1,..,k$ multivariaten ($p>1$) Beobachtungen, ob die apriori-Klassifikation der Beobachtungseinheiten Unterschiede in den Gruppen wiedergibt (Alternative A), oder ob die Beobachtungen eine Zufallsaufteilung der $n=\sum_{i=1}^{k} n_i$ Einheiten darstellen (Hypothese H). Für die Beobachtungen $y_{ij}=(y_{ij1},\ldots,y_{ijp})'$ - $i=1,..,k; j=1,..,n_i$ werden keine Annahmen über die zugrundeliegende(n) Verteilung(en), aus der (denen) sie stammen sollen, gemacht. Die Annahme (i) einer endlichen Population mit diesen n Beobachtungen ist ausreichend; die zuvor übliche Annahme (ii) der $\tilde{S}_n$ -Invarianz (vgl. Abschnitt 4.4 (iii) der gemeinsamen Fehlerverteilungsfunktion sowie der Zufallsvariablen der Beobachtungsvektoren selbst bei Gültigkeit der Nullhypothese ist ebenfalls hinreichend zur Anwendung der von Mielke und Mitarbeitern vorgeschlagenen Klasse von sog. Multi-Response Permutation Procedures, bei denen es sich um exakte Randomisierungs/Permutationtests handelt. Ein weiterer Fall, die Annahme (iii): die $k \geq 1$ Gruppen sind k Zufallsstichproben aus einer endlichen Grundgesamtheit von $N>n$ Einheiten, wird ebenfalls expliziert von Mielke diskutiert. In dieser Situation soll

die Hypothese überprüft werden, ob die k Gruppen identisch (H) oder verschieden (A) von der endlichen Population sind. Im folgenden wird, wenn nicht anders erwähnt, der Fall (ii) betrachtet.

Ausgangspunkt zur Konstruktion einer Teststatistik ist irgendeine Norm $\Delta(Y_{ij},Y_{i'j'}) = ||Y_{ij}-Y_{i'j'}||$; $i,i'=1,..,k$; $j,j'=1,..,n_i$ bzw. $n_{i'}$.
Üblicherweise werden jedoch nur Euklidische Distanzen (oder deren Quadrate) betrachtet:

(4.66) $$\Delta(Y_{ij},Y_{i'j'}) = \left(\sum_{l=1}^{p} Y_{ijl}-Y_{i'j'l}\right)^{v/2} \quad \text{mit } v=1 \text{ (oder 2)} .$$

Als Teststatistik werden die Mittelwerte aller Distanzen von Paaren von Elementen innerhalb der k Gruppen gewählt (Berry & Mielke, 1983a):

(4.67) $$\delta(Y) = \sum_{i=1}^{k} n_i/n \cdot \binom{n_i}{2}^{-1} \sum_{j<j'}^{n_i} \Delta(Y_{ij},Y_{ij'})$$
$$= 2/n \sum_{i=1}^{k} 1/(n_i-1) \sum_{j<j'}^{n_i} \Delta(Y_{ij},Y_{ij'}) \quad .$$

Bei Gültigkeit der Nullhypothese H sind alle $m:=n!/\prod_{i=1}^{k} n_i!$ Zuordnungen zu den k Gruppen gleichwahrscheinlich mit Wahrscheinlichkeit $\tilde{P}_n = (n!/\prod_{i=1}^{k} n_i!)^{-1}$.

Zur Durchführung eines exakten Permutationstests kann man wie zuvor wiederholt dargestellt vorgehen. Zu beachten ist nur, daß in diesem Fall 'kleine' Werte der Teststatistik gegen die Beibehaltung der Nullhypothese sprechen. H wird verworfen, falls für (ein erzielbares) $\alpha, 0<\alpha<1$, gilt:

(4.68) $$p^{\leq} = |\{\pi \in \tilde{S}_n \mid \delta(\pi Y) \leq \delta(Y)\}|/m \leq \alpha \quad .$$

Denn wenn die Gruppierungen der Beobachtungen 'echte' Unterschiede in den Verteilungen der Werte in den k Gruppen widerspiegelt, sollten bei Permutationen der Beobachtungen häufiger größere mittlere Abstände je Gruppe resultieren.
Ein exakter Permutationstest ist jedoch für diese Art von Statistik besonders rechenaufwendig. Denn pro neuer Permutation müssen $\sum_{i=1}^{k} \binom{n_i}{2}$ Abstände gegenüber k Summationen von je n_i Vektoren im Fall einer MANOVA-Teststatistik gebildet werden. Deshalb wird von den Testautoren versucht, die exakte Verteilungsfunktion von $\delta(Y)$ aus (4.67) mit Hilfe der ersten drei Momente der Permutationsverteilung von $\delta(Y)$ zu approximieren. Sei:

(4.69) $\mu_\delta := E_{P_n}(\delta(Y)) = 1/m \sum_{j=1}^{m} \delta_j$

- j=1,..,m indiziert alle $n!/\Pi n_i!$ unter H zulässigen Permutationen der Beobachtungen -

$$\sigma_\delta^2 := Var_{P_n}(\delta(Y)) = 1/m \sum_{j=1}^{m} \delta_j^2 - \mu_\delta^2$$

$$\gamma_\delta := 1/m \cdot \sigma_\delta^3 \left(\sum_{j=1}^{m} \delta_j^3 - 3\mu_\delta\sigma_\delta^2 - \mu_\delta^3 \right) .$$

Formeln zur Berechnung der drei Momente sind z.B. in Mielke et al. (1976) angegeben. Tracy & Tajuddin (1985, 1986) bestimmen sogar die ersten vier Momente von δ (vgl. auch Siemiatycki, 1978).
Die standardisierte Teststatistik

(4.70) $T_\delta := (\delta - \mu_\delta)/\sigma_\delta$

kann durch eine Pearson-Typ III Verteilungsfunktion, in die die Schiefe γ_δ eingeht, approximiert werden (Mielke et al, 1976, S. 1414f). Damit ist es ohne Durchführung von Permutationen möglich, einen approximativen p-Wert auszurechnen (Berry & Mielke, 1983a). Man kommt also zu dem interessanten

Ergebnis, daß die asymptotische Verteilung von δ für $n\to\infty$, $n_i/n\to\lambda_i\in(0,1)$ nicht eine Normalverteilung ist sondern eine schiefe Verteilung:

$(k-1)-T_\delta(2(k-1))^{1/2}$ strebt gegen eine χ^2-Verteilung mit $k-1$ Freiheitsgraden (Brockwell & Mielke, 1982; Robinson, 1983). Berry & Mielke (1983a) haben neben einem Algorithmus für den exakten Test (Berry, 1982; Berry & Mielke, 1982, 1984) ein Programm zur Durchführung des approximativen Tests publiziert (vgl. Abschnitt 8.1).

Für eindimensionale Beobachtungen und $\Delta=(Y_{ij}-Y_{ij'})^2$ für alle $i=1,..,k$ hat

$$(4.71)\quad \delta_1(Y) = \sum_{i=1}^{k}(n_i-1)/(n-k)\cdot\binom{n_i}{2}^{-1}\sum_{j<j'}^{n_i}\Delta(Y_{ij},Y_{ij'})$$

einen einfachen Zusammenhang zur F-Statistik der ANOVA (Mielke, Berry & Medina, 1982; Berry & Mielke, 1983b). Es gilt:

$$(4.72)\quad \delta_1(Y) = 2\sum_{ij}(Y_{ij}-\bar{Y}_{..})^2/(n-k+(k-1)F(Y)) \quad .$$

Ein Rangtest für diesen Spezialfall $p=1$ ist in Mielke, Berry, Brockwell & Williams (1981) dargestellt. Mielke & Sen (1981) behandeln Rangtests für $k\geq 2$ in allgemeiner Form.

Für den Fall einer endlichen Population mit N Elementen (Annahme (iii)) und $k\geq 2$ Gruppen mit $n=\sum_{i=1}^{k} n_i$ Elementen $n<N$ kann gezeigt werden, daß

$$(4.73)\quad \delta_N(Y) = \Big(\sum_{i=1}^{k}\binom{n_i}{2}\Big)^{-1}\sum_{i=1}^{k}\sum_{j<j'}^{n_i}\Delta(Y_{ij},Y_{ij'})$$

Mielke et al., 1976; Mielke, 1978, 1979b)

wie auch

$$\delta'_N(Y) = \sum_{i=1}^{k} n_i/n\cdot\binom{n_i}{2}^{-1}\sum_{j<j'}^{n_i}\Delta(Y_{ij},Y_{ij'})$$ (Mielke, 1979a)

mit Permutationsmaß $P_N = N! / \prod_{i=1}^{k+1} n_i!$, d.h. Vertauschung über alle N Elemente der endlichen Population ($n_{k+1} = N-n$) beide asymptotisch nur dann eine schiefe Verteilung besitzen, falls N>n. Je näher N an n liegt, umso langsamer konvergiert der Schiefeparameter gegen Null (O'Reilley & Mielke, 1980). Mielke (1978, 1979a) zeigt, daß δ'_N (aber damit auch $\delta, \delta_1, \delta_N$) alle Spezialfälle einer allgemeinen Permutationsstatistik von Mantel (1967)

$$\sum C_{I_1,I_2} \Delta_{I_1,I_2}$$

sind mit Summation über alle Permutationen geordneter Paare von nicht-identischen Indizes $I_1, I_2 \in \{1,..,N\}$.

Für das p-dimensionale verbundene k-Stichprobenproblem (vgl. Abschnitt 4.5 (iv)) können ebenfalls MRPP-Statistiken verwendet werden (Mielke & Iyer, 1982; Iyer et al., 1983). Diese sog. MRBP (Multi Response Randomized Block Permutation Procedures)-Teststatistik hat nach Eliminierung der 'störenden' Blockeffekte durch Subtraktion der Blockmittel folgende Gestalt:

(4.74) $$\delta_o(T) = (k\binom{r}{2})^{-1} \sum_{i=1}^{k} \sum_{j<j'}^{r} \Delta(T_{ij}, T_{ij'}) \quad ,$$

mit T_{ij} aus (4.30). Mielke & Iyer geben lediglich eine heuristische Begründung für diese Transformation. Sie soll verhindern, daß große Blockunterschiede Effekte des k-stufigen Faktors verdecken konnen.
Unter der üblichen Nullhypothese haben alle $(k!)^r$ Intrablockpermutationen dieselbe Wahrscheinlichkeit $P_n = 1/(k!)^r$. Die Nullhypothese wird bei Durchführung des exakten Permutationstests wiederum verworfen, falls (bei erzielbarem α, $0<\alpha<1$):

(4.75) $$p^{\leq} = |\{\pi \in \times_{j=1}^{r} S_k \,|\, \delta_o(\pi T) \leq \delta_o(T)\}| / (k!)^r \leq \alpha$$ ist.

Erneut ist der Rechenaufwand sehr groß, denn pro Permutation müßten $k\cdot\binom{r}{2}$ Euklidische Distanzen (v=1) berechnet werden. Mielke & Iyer (1982) berechnen wieder die ersten drei exakten Permutationsmomente von δ_o unter dem Permutationsmaß P_n und approximieren wie im unverbundenen k-Stichprobenproblem die exakte Verteilung der standardisierten Statistik $T_o=(\delta_o-\mu_{\delta_o})/\sigma_{\delta_o}$ durch eine Pearson-III Verteilung, die die Schiefe von δ_o berücksichtigt (Computer-Programm in Iyer et al., 1983).

Für den univariaten Fall besteht wieder ein funktionaler Zusammenhang zur F-Statistik der ANOVA (Mielke & Iyer, 1982).

Falls Replikationen pro Block vorhanden sind ($n_{ij}\geq 2$, vgl. Abschnitt 4.5 (v)), werden für den MRBP-Ansatz zwei verschiedene Vorgehensweisen vorgeschlagen. Im sog. 'cell means approach' sind zuerst die Zellmittelwerte über die Replikationen zu bilden; diese werden dann aligniert und man verfährt weiter wie im Fall $n_{ij}=1$. Der sog. 'align and collapse approach', in dem die individuellen Beobachtungen um die Blockmittel aligniert und dann die p-dimensionalen Vektoren je Faktorstufe wie die Replikationen unter einer Faktorstufe in pCR-k behandelt werden unter Benutzung der Teststatistik δ aus (4.67), verletzt die Blockstruktur des Versuchsplanes. Zwar sind die geschätzten Blockeffekte subtrahiert worden, die Permutationsinvarianzstruktur der alignierten Fehlervektoren bleibt aber weiterhin $\times_{i=1}^{k} S_{n_{i.}}$, sodaß die Momente von δ unter dem Permutationsmaß $1/\Pi n_{i.}!$ und nicht unter $1/(n!/\Pi n_i!)$ wie in Mielke et al., (1976) für das unverbundene k-Stichprobenproblem bestimmt werden müssen.

Das verbundene 2-Stichprobenproblem mit p=1 wird von Mielke & Berry (1983) und Brockwell & Mielke (1984) mit einem Computerprogramm von Berry & Mielke (1985) behandelt.

4.7.2 Verallgemeinerung des MRPP-Ansatzes auf mehrfaktorielle multivariate Versuchspläne

Wie im pRB-k Versuchsplan für die Beseitigung der Blockeffekte schon demonstriert, kann folgendes heuristische Konstruktionsprinzip für MRPP-Statistiken in mehrfaktoriellen, multi- wie auch univariaten Versuchsplänen angewendet werden:

(1) Man transformiert die ursprünglichen Zufallsvektoren wie für die Invarianz-Permutationstests, um 'störende' Effekte beim Testen der jeweiligen Nullhypothese zu eliminieren (vgl. (3.6)).

(2) Man berücksichtigt wie zuvor die Permutationsinvarianzstruktur der Verteilungsfunktion des transformierten Beobachtungsvektors.

(3) Anstelle der Abbildung in den zum linearen Unterraum der linearen Hypothese orthogonalen Raum und der Bildung der quadrierten euklidischen Länge dieser Projektion (vgl. (3.16)), d.h. anstelle der Bildung der Summe über alle Indizes der nach Transformation verbliebenen Parameter von über alle restlichen Indizes gemittelten und dann quadrierten transformierten Beobachtungen summiert man nun wieder über alle Indizes der verbliebenen Parameter, diesmal jedoch die mittleren (quadrierten) Distanzen zwischen allen Paaren von transformierten Beobachtungen, über die bei

den Invarianz-Permutationstests gemittelt wird. Es werden also alle Distanzen von Paaren transformierter Beobachtungen summiert, die bezüglich des transformierten Parametervektors identisch, d.h. Replikationen, sind.

Beispiel: Im multivariaten Split-Plot Design mit Modellgleichung (4.19) haben die transformierten Beobachtungsvektoren

$T_{ihj}=(T_{ihj1},\ldots,T_{ihjp})'$ die Gestalt

$$T_{ihj} = Y_{ihj}-\bar{Y}_{ih.}-\bar{Y}_{..j}+\bar{Y}_{...} \quad .$$

Die Verteilungsfunktion von $T=(T'_{111},\ldots,T'_{rn_rc})'$ ist $\tilde{S}_{n_.} \otimes \tilde{S}_c$-invariant. Für die Komponenten des transformierten Parametervektors gilt:

$$(\hat{\tilde{K}}\theta)_{ihj}=(\alpha\beta)_{ij} \qquad \text{für alle } h=1,..,n_i \text{ und jedes } i=1,..,r;\ j=1,..,c.$$

Damit ist die mittlere Distanz aller Paare von transformierten Beobachtungen je Plotstufe i und Splitfaktorstufe j über alle Replikationen (Blöcke) zu bilden. Diese mittleren Distanzen sind dann nach Gewichtung um die unterschiedlichen Stichprobenumfänge über alle Faktorstufenkombinationen (i,j) zu summieren:

$$(4.76) \qquad \delta(T) = \sum_{i=1}^{r} n_i/n_. \cdot \sum_{j=1}^{c} \binom{n_i}{2}^{-1} \sum_{h<h'}^{n_i} \Delta(T_{ihj},T_{ih'j})$$

Zur Durchführung des exakten Permutationstests sind alle $n_./\prod_i n_i!$ Permutationen der pc-dimensionalen $T_{ih}=(T'_{ih1},\ldots,T'_{ihc})'$ auszuführen, $\delta(\pi T)$ zu berechnen und

$$p^{\leq}(T) = |\{\pi \in S_{n_\cdot} \otimes S_c \mid \delta(\pi T) \leq \delta(T)\}| \cdot (\prod_i n_i!/n!)$$

zu ermitteln. Die Permutationen von Split-Stufen bewirken lediglich Umindizierungen in j und lassen δ invariant.

Zur Approximation der exakten Verteilung von $\delta(T)$ sind erneut die ersten drei Permutationsmomente von $\delta(T)$ zu bestimmen.

5. MULTIPLE PERMUTATIONSTESTS

Üblicherweise befindet man sich bei der Durchführung experimenteller Untersuchungen in der Situation, aufgrund der erhobenen uni- oder multivariaten Beobachtungen mehrere Entscheidungen treffen zu wollen. Werden mehrere, voneinander abhängige oder unabhängige statistische Tests ausgeführt, muß u.a. Vorsorge getroffen werden, daß nicht das α -Niveau für die statistische Gesamtaussage bezüglich aller Tests zusammengenommen unkontrolliert bleibt. Es ist also wichtig, die Wahrscheinlichkeit, mindestens eine oder irgendeine der einzelnen Nullhypothesen falschlicherweise zu verwerfen, bei α, $0<\alpha<1$, kontrollieren zu können. Äquivalent zu dieser Forderung soll die Wahrscheinlichkeit, alle wahren einzelnen Nullhypothesen nicht zu verwerfen, nicht kleiner als $1-\alpha$ sein (vgl. Sonnemann, 1981a, b, 1982; Miller, 1981).

In diesem Abschnitt sollen multiple Permutationstests dargestellt werden, mit denen das α-Niveau möglichst gut ausgeschöpft wird, und für die die Tests bezüglich einzelner Teilhypothesen möglichst große Teststärke besitzen. Die in den Sozialwissenschaften in den letzten Jahren vermehrt diskutierten Probleme der möglichst strengen Prüfung wissenschaftlicher Hypothesen (Bredenkamp, 1972, 1980, 1983, 1984; Hager & Westermann, 1983a, b, c, 1984; Meehl, 1967) und die Konsequenzen für die Inferenzstatistik im Sinne von Neyman & Pearson (vgl. Lehmann, 1959) werden erst in Abschnitt 7 berücksichtigt. Dazu sind Überlegungen zur Bestimmung der Teststärke bei Permutationstests und die Schätzung von Effektstärken notwendig (vgl. Abschnitt 6).

Bei multivariaten Testproblemen in linearen Multiparametermodellen sind über eine 'globale' Nullhypothese bezuglich der p-dimensionalen Modellparameter hinaus zwei Arten von Teilhypothesen von besonderem Interesse: Hypothesen über Teilmengen (und Linearkombinationen) der p abhängigen Variablen und/oder Teilräume (und lineare Kontraste) des gesamten p-dimensionalen Parameterraumes. Bei parametrischen multiplen (oder simultanen) Testprozeduren (vgl. z.B. Gabriel, 1968, 1969a, b) können beide Arten von Fragestellungen gleichartig behandelt werden. Für Permutationstests ergeben sich Probleme, wenn Teilhypothesen bezüglich der Parameter getestet werden sollen und man die Permutationsinvarianzstrukturen unter Gültigkeit der globalen Nullhypothese benutzen will (vgl. Abschnitt 5.1.4; Petrondas & Gabriel, 1983).

5.1 Bekannte Verfahren für uni- und multivariate multiple Permutationstests

5.1.1 Verfahren, die auf Zweistichproben-Permutationstests basieren

Edgington (1980, S. 86f) schlägt für das unverbundene k-Stichprobenproblem vor, für alle k(k-1)/2 Paarvergleiche zwischen je zwei Stichproben jeweils einen Permutationstest fur das unverbundene univariate Zweistichprobenproblem auszuführen. Dabei ist es unerheblich, ob gleiche oder ungleiche Stichprobenumfänge vorliegen. Der experimentweise Fehler 1. Art α_{EW} (vgl. Kirk, 1968, Kap. 3) wird zu α festgesetzt. Bei jedem einzelnen Paarvergleich wird ein Niveau

von $\alpha/(k(k-1)/2)$ festgelegt. Äquivalent kann man jeden der $k(k-1)/2$ p-Werte mit $k(k-1)/2$ multiplizieren und alle 'adjustierten' p-Werte gegen α_{EW} vergleichen. Die Bonferroni-Ungleichung (vgl. Hunter, 1976) sichert, daß α_{EW} eine obere Schranke für den Gesamtfehler 1. Art ist.

Diese Vorgehensweise ist nicht auf das k-Stichprobenproblem beschränkt, sondern sie kann stets eingeschlagen werden, wenn k Stufen eines Faktors mit oder ohne wiederholte Messungen miteinander verglichen werden sollen. Pyhel (1980b) schlägt das Verfahren explizit für den Vergleich der k Faktorstufen bei Lateinischen Quadraten vor, erwähnt jedoch nur, daß die Zweistichproben-Permutationstests 'angemessen' zu wählen seien. Für das Lateinische Quadrat bedeutet das, daß aus den nach Transformation der Beobachtungen zulässigen Zeilen- und Spaltenpermutationen (vgl. 3.1.5 (ix)) jeweils für einen paarweisen Vergleich nur diejenigen ausgewählt werden dürfen, in denen nur transformierte Beobachtungen vertauscht werden, die unter den beiden jeweils betrachteten Faktorstufen gemacht wurden (vgl. Abschnitt 5.4).

Im verbundenen k-Stichprobenproblem sind für die Paarvergleiche Permutationstests für das verbundene Zweistichprobenproblem einzusetzen. Die beschriebene Vorgehensweise ist nicht auf paarweise Vergleiche beschränkt. Sogenannte 'many-one' Vergleiche, d.h. paarweise Vergleiche zwischen je einer von mehreren experimentellen Bedingungen und einer Kontrollbedingung sind ebenfalls entsprechend ausführbar.

Anstelle der Bonferroni-Prozedur kann stets ihre gleichmäßige Verbesserung, die Bonferroni-Holm-Prozedur (Holm, 1979) verwendet werden (Pyhel & Willmes, 1981; Willmes,

1982). Diese schöpft das Gesamtniveau α_{EW} besser aus. Für den Fall aller paarweisen Vergleiche von Faktorstufen lassen sich die Niveaus dieses Verfahrens noch weiter verschärfen (Hommel, 1984, 1985; Shaffer, 1984, 1986).

5.1.2 Verfahren, die auf den bei Gültigkeit der globalen Nullhypothese zulässigen Permutationen basieren

Miller (1981, S. 179ff) behandelt eine direkte Übertragung von Tukey's T-Methode für den Fall k unabhängiger Stichproben mit identischen Stichprobenumfängen und Verteilungen, die sich lediglich in ihrem Lageparameter unterscheiden. Als Teststatistik wird

$$(5.1) \quad Q(Y) = \max_{i,i'} \{|\bar{Y}_{i.} - \bar{Y}_{i'.}|\} \; ; \quad \bar{Y}_{i.} = 1/n \sum_{j=1}^{n} Y_{ij} \; - \; i=1,..,k \; -$$

die Spannweise (range) der Stichproben-/Faktorstufenmittel betrachtet.

Bei Gültigkeit der globalen Nullhypothese sind alle (kn)! Permutationen der Beobachtungen gleichwahrscheinlich. Wie zuvor dargestellt, können nur m:=(kn)!/kn!k! von ihnen zu unterschiedlichen Teststatistikwerten führen.

Zur Durchfuhrung aller paarweisen Vergleiche wird die exakte Verteilungsfunktion von Q(Y) durch Berechnung von (5.1) für alle zulässigen Permutationen bestimmt. Für das globale Fehlerniveau $\alpha_{EW} = \alpha$ wird das $m(1-\alpha)+1$-te Quantil $Q_{m,1-\alpha}$ dieser diskreten Verteilungsfunktion bestimmt. α sei so gewählt, daß $m(1-\alpha)$ eine ganze Zahl und ein 'erreichbares' Niveau ist; ansonsten wird $[m(1-\alpha)]$ gewählt.

Alle $k(k-1)/2$ absoluten Differenzen von Stichprobenmitteln werden mit demselben Schwellenwert $Q_{m,1-\alpha}$ verglichen und zwei Faktorstufen i, i' werden als signifikant verschieden erklärt, wenn

$$|\bar{y}_{i.}-\bar{y}_{i'.}| \geq Q_{m,1-\alpha}(y) \quad , \quad y\in R^{nk} \text{ und } y=(y_{11},..,y_{kn})' \ .$$

Auch alle Spannweisen von $2\leq l\leq k$ Stichproben

$$q_i = \max_{i\in J} \bar{y}_{j.} - \min_{i\in J} \bar{y}_{j.} \quad \text{mit } i\in J\subset\{1,..,k\} \ , \ |J|=l$$

können gegen diesen Schwellenwert $Q_{m,1-\alpha}$ verglichen werden. Shuster & Boyett (1979) haben den Ansatz auf das verbundene k-Stichprobenproblem mit Wiederholungen mit $n_{il}>1$, $i=1,..,r$ Blocke, $l=1,..,k$ Faktorstufen, verallgemeinert (Modellgleichung siehe (3.42)). Bei Gültigkeit der globalen Nullhypothese sind alle $\prod_{i=1}^{r} n_i!$ Intrablockpermutationen gleichwahrscheinlich.

Man bildet die exakte Verteilung von

(5.2) $$Q(Y) = \max_{l,l'} \{|Y_{..l}-Y_{..l'}|\} \ , \ Y_{..l}=\sum_{i=1}^{r} n_{i.}\bar{y}_{i.l} \ - \ l=1,..,k \ -$$

unter allen $m=\prod_i(n_{i.}!/\prod_l n_{il}!)$ Permutationen, die zu verschiedenen Teststatistikwerten führen können. Wie zuvor wird $Q_{m,1-\alpha}$ bestimmt und für alle Paare l, l' von Faktorstufenmitteln wird $|y_{..l}-y_{..l'}|$ mit $Q_{m,1-\alpha}(y)$ verglichen. Erneut können auch Spannweiten bezüglich einer Untermenge von Stichproben gegen denselben obigen Schwellenwert verglichen werden.

Mit demselben Ansatz zur Konstruktion multipler Permutationstests ist es auch möglich, alle (einseitigen) Vergleiche gegen eine Kontrolle ($l=1$) zu bilden. Dazu wird das entsprechende obere Quantil der exakten Verteilung von

(5.3) $$Q(Y) = Y_{..1} - \min_{l\in\{2,..,k\}} Y_{..l}$$

gebildet und jede der k-1 Differenzen $y_{..1} - y_{..1}$, $1=2,..,k$, gegen $Q_{m,1-\alpha}(y)$ verglichen.

Für $r=1$ und $n_{i1}=n_1$ hat man den Spezialfall des unverbundenen k-Stichprobenproblems mit ungleichen Stichprobenumfängen; für $r>1$ und $n_{i1} = 1$ für alle i,l den RB-k Plan.
Pyhel (1980b) schlägt das Verfahren auch für Lateinische Quadrate vor. Generell können nach eventueller Transformation der Ursprungsdaten zur Beseitigung 'störender' Effekte Vergleiche für die k Stufen eines Faktors ausgeführt werden, wenn man die exakte Verteilung von Q(TY) unter den nach Transformation zulässigen Permutationen bestimmt.

Foutz (1984) wählt einen anderen Zugang zum Vergleich aller Paare von k Faktorstufen im unverbundenen k-Stichprobenproblem. Er betrachtet den $r=k(k-1)/2$-dimensionalen Vektor F von F-Statistiken für Paarvergleiche:

$$F(Y) = (F_{12},..,F_{ii'},..,F_{k-1,k})'$$

mit

(5.4) $$F_{ii'}(Y) = (\bar{Y}_{i.}-\bar{Y}_{i'.})^2/(S^2(1/n_i + 1/n_{i'})) \; ; \; i\neq i'=1,..,k$$

und

$$S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2/(n_{.}-k) \quad ; \; n_{.} = \sum_{i=1}^{k} n_i \; .$$

Die Y_{ij} seien verteilt nach $G_i=G(x-\mu_i)$ (vgl. Abschnitt 3.1.5 (iii).

Unter Gültigkeit der Familie H von Nullhypothesen $H=\{H_o^{ii'}: \mu_i=\mu_{i'}\}$ sind alle $F(\pi Y)$ gleichwahrscheinlich mit $\pi\in S_{n_.}$; von diesen können wiederum nur $N:=n_.!/\Pi n_i!$ zu verschiedenen Werten $F^1,..,F^N$ von F(Y) führen.
Seien o.B.d.A. alle Werte $F^1_{ii'},..,F^N_{ii'}$ verschieden für jedes Paar i,i'. Zur Durchführung des exakten Permutations-

tests verwendet man folgenden Algorithmus zur Erzeugung des oberen α-ten Anteils der gemeinsamen Verteilung der $F^1,..F^N$, $0<\alpha<1$, (Foutz, 1981; Foutz, Jensen, & Anderson, 1985):

(i) In einer (o.B.d.A.) zufälligen Reihenfolge werden insgesamt $M:=[\alpha N]$-mal alle geordneten Index-Paare (i,i'), $i<i' \in \{1,..,k\}$ jeweils M/r-mal betrachtet. Seien sie mit $h(1),..h(M)$ bezeichnet und sei o.B.d.A. α so gewählt, daß M eine ganze Zahl und M/r ohne Rest teilbar sei.

(ii) Der größte F-Wert unter den N Werten in dem $h(1)$-ten Paarvergleich wird ausgewählt (und gespeichert); alle weiteren Komponenten dieses F-Vektors werden nicht weiter berücksichtigt.

(iii) Dann wird der größte F-Wert unter den verbleibenden N-1 Werten in dem $h(2)$-ten Paarvergleich ermittelt und ebenfalls gespeichert.

(iv) Nach $l-1$ Schritten sind $l-1$ $F_{h(\mu)}$-Werte, $\mu=1,..,l-1$, ausgewählt, die je Paarvergleichskomponente mit steigendem l stets kleiner werden. Im $h(l)$-ten Paarvergleich wird unter den $N-(l-1)$ verbliebenen Werten wieder der größte gesucht.

(v) Nach M Schritten sind die kritischen Werte $c_1(Y),..,c_r(Y)$ für die einzelnen Paarvergleichskomponenten gerade die jeweils kleinsten ausgewählten M/r-ten $F_{ii'}$ Statistikwerte.

Unter Gültigkeit aller Hypothesen der Familie von Hypothesen H ist damit die Wahrscheinlichkeit, daß wenigstens eine der Paarvergleichshypothesen verworfen wird $\alpha = M/N$.

Foutz definiert auch eine Marginal-Fehlerwahrscheinlichkeit 1. Art $\alpha_{ii'}$, als die Wahrscheinlichkeit, die Nullhypothese $H_o^{ii'}:\mu_i=\mu_{i'}$ zu verwerfen, wenn alle Hypothesen der Familie gültig sind. Der zugehörige p-Wert ist:

(5.5) $p_{ii'}(\underline{y}) = 1/N \cdot |\{\pi \in S_n : F_{h(l)}(\pi \underline{y}) \geq c_{ii'}(\underline{y}),\ l=1,..,M\}|$

mit M/r Werten $h(l)=(i,i')$.

Damit ist ein allgemeines Prinzip der Konstruktion von Permutationstests zum simultanen Testen einer Familie von Hypothesen vorgestellt worden. Es muß lediglich die Bedingung erfüllt sein, daß bei Gültigkeit aller Hypothesen alle zulässigen Permutationen der Beobachtungen gleichwahrscheinlich sind. Im verbundenen k-Stichprobenproblem z.B. sind bei Gültigkeit aller $H_o^{ii'}$ von zuvor alle $(k!)^n$ Intrablockvertauschungen der um die Blockmittel 'bereinigten' Beobachtungen gleichwahrscheinlich.

Das Gesamtniveau α hängt nicht von der Zufälligkeit der Reihenfolge der Index-Paare in (i) und nicht von der Wahl $\alpha_{ii'}=(M/R)/N$ ab, lediglich die Marginalniveaus werden davon beeinflußt. Deshalb wird i.a. wie in (i) vorgegangen.

Falls N sehr groß ist, kann auch eine sog. Monte-Carlo Lösung ('randomized permutation test', vgl. Abschnitt 8.2), in der nur eine Zufallsstichprobe vom Umfang $N^* \ll N$ aller zulässigen Permutationen betrachtet wird, zum Testen verwendet werden. Die identische Permutation sei stets in diese Stichprobe eingeschlossen.

Anstelle der Teststatistiken (5.4) genügt es, für das unverbundene k-Stichprobenproblem

(5.6) $\hat{S}_{ii'}(\underline{Y}) = |\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}|$ $i \neq i' \in \{1,..,k\}$ oder $(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot})^2$

zu betrachten.

Als Hypothesen der Hypothesenfamilie können z.B. auch alle Hypothesen über Vergleiche gegen eine Kontrolle betrachtet werden mit:

(5.7) $S_i(Y)=\bar{Y}_{i.}-\bar{Y}_{k.}$, i=1,..,k-1 und k als Kontrollbedingung. Es sei hier bereits kritisch angemerkt, daß das Verfahren von Foutz lediglich das globale, aber nicht das multiple α-Niveau kontrolliert.

5.1.3 Gabriel's simultane Testprozedur (STP) für simultane Permutationstests

Die in Abschnitt 5.1.2 vorgestellten Permutationstests von Miller bilden einen Spezialfall des allgemeinen Verfahrens für simultane (Permutations-) Tests, das Gabriel (1979) jedoch nur kurz skizziert. Betrachtet wird eine Familie von Hypothesen (für eine detaillierte Definition vgl. Abschnitt 5.2!) $\{H_i\}$, $i\in I$ (I eine o.B.d.A. endliche Indexmenge) sowie deren Durchschnitt $H:=\bigcap_{i\in I} H_i$ (Gabriel, 1969a). Zu jeder Hypothese H_i gehöre eine Teststatistik S_i. Dann ist $S=\max_{i\in I} S_i$ eine geeignete Teststatistik für H. Falls die S_i bei Gültigkeit von H_i nicht dieselbe Verteilungsfunktion besitzen, kann man auch die jeweils zugehörigen p-Werte $p_i=P_{H_i}(S_i\geq s_i)$ und $p=\min_{i\in I} p_i$ als Teststatistik wählen. Dann werden die S_i (bzw. p_i) und S (bzw. p) für alle unter H zulässigen Permutationen der Beobachtungen bestimmt. Das obere (bzw. untere) erreichbare α-Quantil der Verteilungsfunktionen von S (bzw. p) wird bestimmt und H sowie (alle) H_i verworfen, falls deren Teststatistikwerte für die (unpermutierten) Beobachtungen größer oder gleich (bzw. kleiner oder gleich) dem berechneten α-Quantil sind.

Für das p-variate unverbundene nichtparametrische k-Stichprobenproblem (vgl. Abschnitt 4.5 (iii)) ist die Familie aller $(2^k - k - 1)(2^p - 1)$ Hypothesen über die Gleichheit von Teilmengen aller k Lageparameter-Teilvektoren der p-dimensionalen Vektoren von Interesse (vgl. Gabriel, 1968, 1969b). Willmes (1981a) hat für diese Hypothesenfamilie simultane Permutationstests vorgeschlagen.

Allgemein haben die 'simultanen' Permutationstest-Statistiken nicht nur für das unverbundene k-Stichprobenproblem folgende Gestalt (mit den Bezeichnungen aus Abschnitt 4):

$$\psi^{G,V}(KY) = \psi(((\underline{S}_e^V + \underline{S}_h^{G,V})(KY))^- \cdot \underline{S}_h^{G,V}(KY)) \qquad (5.8)$$

ψ bezeichne eine monotone Funktion der Eigenwerte des Ausdruckes in Klammern, $G \subseteq \{1,..,k\}$; $V \subseteq \{1,..,p\}$; K sei die Transformationsmatrix zur Eliminierung aller störenden Effekte.

Die Indizierungen mit G und V sollen anzeigen, daß jeweils nur die in der (Teil-) Hypothese $H^{G,V}$ betrachteten Faktorstufen und abhängigen Variablen zur Bildung der $|V| \times |V|$-dimensionalen Effekt- und Fehlermatrizen berücksichtigt werden.

Die Teststatistiken sind (wie in der MANOVA) monoton, d.h.

$$\psi^{G',V'}(KY) \leq \psi^{G,V}(KY) \quad , \text{ falls } G' \subseteq G,\ V' \subseteq V \ , \qquad (5.9)$$

denn in dem Beweis der Monotonieeigenschaften (vgl. Gabriel, 1968, Appendix A, B) werden lediglich algebraische Argumente verwendet. Die Gleichheit in (5.9) gilt (fast sicher) nur, wenn ein Maximalwurzelkriterium (vgl. (4.18)) verwendet wird.

Alle Teststatistikwerte für Teilhypothesen werden gegen denselben oberen bzw. unteren α-Schwellenwert (je nach Wahl der Teststatistik) der exakten Permutationsverteilung von

(5.8) verglichen. Dieser wird anhand der unter der globalen Nullhypothese $H^{G,V}$ mit $G=\{1,..,k\}$ und $V=\{1,..,p\}$ zulässigen Permutationen bestimmt.

5.1.4 Inadäquatheit gewisser simultaner Testprozeduren für Permutationstests

Simultane Testprozeduren (STP) sind dadurch gekennzeichnet, daß ein gemeinsamer kritischer Wert für alle Tests von (Teil-) Hypothesen der Hypothesenfamilie verwendet wird (Gabriel, 1969a, Abschnitt 3). Das nominelle α-Niveau der STP ist gleich der Wahrscheinlichkeit, daß bei Gültigkeit der Durchschnittshypothese über alle Hypothesen der Hypothesenfamilie die Teststatistik zum Testen dieser Durchschnittshypothese Werte größer (oder gleich) dem gemeinsamen kritischen Wert annimmt.
Für Permutationstests zum Vergleich der Einflüsse von k Stufen eines Faktors führt dieses Prinzip zu Problemen. Zur Bestimmung des oberen/unteren α-Quantils - je nachdem ob große/kleine Werte gegen die Durchschnittshypothese (Globalhypothese) sprechen - wird die exakte Verteilungsfunktion der Teststatistik unter allen bei Gültigkeit dieser Hypothese zulassigen Permutationen der Beobachtungen errechnet, wie z.B. in Übertragung der Tukey-Methode in Abschnitt 5.1.2. Petrondas & Gabriel (1981, 1983) zeigen am Beispiel eines Lateinischen Quadrates (LS-3) und des unverbundenen k-Stichprobenproblems, daß bei Berücksichtigung aller unter der globalen Nullhypothese zulässigen Permutationen das nominelle α-Niveau beim Testen von einzelnen Teilhypothesen

uber- oder unterschritten werden kann in Abhängigkeit von Beobachtungen unter anderen Faktorstufen, über die die Teilhypothese keine Aussage macht. Zur Entscheidung über eine Teilhypothese dürfen also, um antikonservative aber auch konservative Testentscheidungen zu vermeiden, nur solche unter der globalen Nullhypothese zulässigen Permutationen der Beobachtungen betrachtet werden, in denen Beobachtungen aus den von der Teilhypothese betroffenen Faktorstufen permutiert werden. Es gibt damit keine generell gültigen Beziehungen zwischen Teststatistiken für Global- und Teilhypothesen, wie z.B. die in Abschnitt 5.1.3 erwähnte Monotonieeigenschaft der MANOVA-Teststatistiken, noch gibt es generelle Beziehungen für die Quantile ihrer Verteilungen. Wie Petrondas & Gabriel selbst feststellen, ist damit der in 5.1.3 dargestellte allgemeine Ansatz von Gabriel (1979), der auch von Willmes (1981a) benutzt wurde, falsch. Das gilt auch für die von Soms (1985) vorgeschlagenen exakten Permutationstests für Paarvergleiche und Vergleiche gegen eine Kontrolle beim unverbundenen k-Stichprobenproblem für binäre Daten. Eine Ausnahme bilden Hypothesenfamilien für multivariate Testprobleme, die nur aus Teilhypothesen bezüglich Komponenten der abhängigen Variablen bestehen. Da die exakte Permutationsverteilung stets fur alle p Komponenten der abhängigen Variablen gemeinsam bestimmt wird, ist die Menge zulässiger Permutationen für die Global- und alle Teilhypothesen stets dieselbe.

Multivariate STPen mit Permutationstests sind somit für Hypothesenfamilien bezüglich der abhängigen Variablen nach dem allgemeinen Prinzip aus Abschnitt 5.1.3 mit Teststatistiken

(5.9) $\psi^V(KY) = \psi(((\underline{S}_e + \underline{S}_h)^V(KY))^- \cdot \underline{S}_h^V(KY))$ und $V \subseteq \{1,..,p\}$

möglich. Sie werden sich jedoch - verglichen mit im folgenden Abschnitt 5.3 dargestellten sequentiell verwerfenden Prozeduren - als weniger teststark erweisen. Zuvor sollen jedoch kurz die in den letzten Jahren entwickelten Konzepte zur Losung multipler Testprobleme aufgezeigt werden, da sie den Rahmen für die im folgenden dargestellten multiplen Permutationstests darstellen.

Der von Petrondas & Gabriel aufgezeigte Fehler gilt auch für exakte simultane Testprozeduren mit (verallgemeinerten) Rangzahlen. Selbst wenn z.B. für Hypothesenfamilien bezüglich aller Paare von Faktorstufen nur eine Rangbildung über alle Beobachtungen der jeweils interessierenden Faktorstufen erfolgt, wird doch die exakte Permutationsverteilung der Rangstatistiken unter allen unter der globalen Nullhypothese zulässigen Vertauschungen der ursprünglichen alignierten oder nicht alignierten Beobachtungen betrachtet. Die Konstellation der Rangzahlen für den individuellen Faktorstufenvergleich ist damit durch Beobachtungen unter anderen Faktorstufen beeinflußt.

5.2 Multiple Tests - Definitionen und theoretische Ansätze

Wie in Abschnitt 2 dargestellt, sind Permutationstests aufgrund ihrer Konstruktion bedingte exakte Tests zum Niveau α. Zum Testen werden Statistiken verwendet, deren Verteilungen bei Gültigkeit einer bestimmten Nullhypothese H invariant sind gegen eine bestimmte Gruppe $\mathcal{G}_N$ von Permutationen der Komponenten des gesamten, ein Experiment be-

schreibenden Beobachtungsvektors X_N oder einer geeigneten Transformierten von X_N. Bei Gultigkeit der betreffenden Nullhypothese besitzen alle Realisationen der geeignet gewählten Teststatistik $S(x_N)$ für die tatsächlichen Beobachtungen sowie deren zulässige M Permutationen die gleiche Wahrscheinlichkeit. Für einen einseitigen (nicht-randomisierten) Permutationstest hat man als Entscheidungsfunktion (vgl (2.5)):

$$(5.10) \qquad \phi(x) = \begin{cases} 1 & p(x) \leq \alpha \\ 0 & p(x) > \alpha \end{cases}$$

$$\text{mit } p(x) = |\{\pi \in \Omega_N \mid S(\pi x) \geq S(x)\}| / M = P_H(S(\pi x) \geq S(x))$$

$$\text{und } P_H(p(x) \leq \alpha) = P_H(\phi = 1) \quad .$$

Im folgenden werden stets anstelle der Teststatistiken ihre Überschreitungswahrscheinlichkeiten, die sog. p-Werte, betrachtet (vgl. z.B. Petrondas & Gabriel, 1983, S. 90). Sie haben gegenüber den Teststatistiken mit i.a. unterschiedlichen Verteilungen den Vorteil, vergleichbar zu sein (vgl. Sonnemann, 1982). Oft sind sie auf dem Intervall [0,1] gleichverteilt.

Falls für ein statistisches Experiment $n \geq 2$ Testprobleme H_o^i gegen H_1^i gegeben sind, handelt es sich um ein multiples Testproblem. Unter einem multiplen Test (einer multiplen Testprozedur) versteht man einen Vektor von Entscheidungsfunktionen $\phi = (\phi_1, .., \phi_n)$, dessen Komponenten jeweils die Werte 0 oder 1 annehmen. Petrondas & Gabriel (1983) haben als erste darauf aufmerksam gemacht, daß zum Testen der einzelnen Hypothesen H_o^i nicht alle $\pi \in \Omega_N$ der Durchschnittshypothese $H_o := \bigcap_{i=1}^{n} H_o^i$ zu berücksichtigen sind, sondern genau die Untergruppe Ω_N^i von Permutationen, die nur diejenigen

Komponenten des Beobachtungsvektors betreffen, auf die sich die einzelne Teilhypothese bezieht. $\mathcal{G}_N^i$ enthält alle Permutationen π^i, die nur die bei Gültigkeit von H_o^i zulässigen Vertauschungen der von H_o^i betroffenen Komponenten von x_N bewirken, ansonsten alle anderen Komponenten von x_N in sich selbst überführen. (Auf eine aufwendigere Schreibweise für die π^i soll hier verzichtet werden.)

Im unverbundenen k-Stichprobenproblem mit $N:=n_\cdot=\sum_{i=1}^{k} n_i$ Beobachtungen zusammengefaßt im Beobachtungsvektor $x_N = (x_{11},..,x_{kn_k})'$ sind beispielsweise bei Gültigkeit der linearen globalen Nullhypothese H_o: $\mu_1 = ... = \mu_k$ alle Permutationen aus $\mathcal{G}_N=S_N$ zulässig. Falls die interessierenden Hypothesen H_o^i alle i=1,..,k(k-1)/2 Paarvergleiche von je 2 Faktorstufen betreffen, sind die $\mathcal{G}_N^i$ gerade die Permutationsuntergruppen, die alle Vertauschungen der Elemente in x_N, die zu dem Stichprobenpaar i=(j,j') gehören, beinhalten.

Sonnemann (1982) stellt dar, daß für ein multiples Testproblem nicht die Kontrolle des globalen Niveaus α, sondern des multiplen Niveaus α von besonderem Interesse ist. Für die Einhaltung des globalen Niveaus α muß bei Gültigkeit der globalen Nullhypothese $H_o=\bigcap_{i=1}^{n} H_o^i$ gelten:

(5.11) $$P\left(\bigcup_{i=1}^{n} (\phi_i=1)\right)\leq\alpha \quad .$$

Für die Einhaltung des multiplen Niveaus α dagegen muß für die Teilhypothesen mit gültiger Nullhypothese

(5.12) $$P\left(\bigcup_{i\in I_o} (\phi_i=1)\right)\leq\alpha \qquad , I_o=\{i \mid H_o^i \text{ gilt}\}$$

erfüllt sein.

Die Wahrscheinlichkeit für einen multiplen Fehler 1. Art, d.h. der Test $\phi=(\phi_1,..,\phi_n)$ entscheidet bei mindestens einem $i\in\{1,..,n\}$ für die Alternative H_1^i, obwohl H_o^i gilt, soll

also kleiner oder gleich α sein.

Aquivalent kann man fordern (vgl. Sonnemann, 1982), daß die Wahrscheinlichkeit, alle wahren Nullhypothesen anzunehmen, mindestens $1 - \alpha$ betragen soll.

Mit obigen Definitionen gilt, daß ein multipler Test zum multiplen Niveau α stets ein Test zum globalen Niveau α ist. Entsprechend zu (5.12) kann man einen multiplen Fehler 2. Art so definieren, daß bei Gültigkeit mindestens einer Alternative H_1^i für diese $i \in \{1,..,n\}$ der Test $\phi_i=0$ entscheidet. In einer multiplen Testprozedur können also sowohl multiple Fehler 1. Art wie 2. Art auftreten.

Zwei weitere wichtige Eigenschaften von multiplen Tests sind Kohärenz und Konsonanz (vgl. Gabriel, 1969a; Sonnemann, 1982). Ein multipler Test heißt kohärent, wenn für alle Nullhypothesen mit der Bedingung H_o^i ist in H_o^j enthalten, d.h. H_o^i impliziert H_o^j ($H_o^i \subseteq H_o^j$) gilt:

Falls H_o^j verworfen wird, gilt dasselbe auch für H_o^i, d.h.

$$\{\phi_j=1\} \subseteq \{\phi_i=1\} \quad .$$

Zur Definition von Konsonanz sei vorausgesetzt, daß die Hypothesenfamilie durchschnittsabgeschlossen ist, d.h. mit je zwei (Null-) Hypothesen gehört auch ihr Durchschnitt zur Hypothesenfamilie, oder äquivalent:

für alle Teilindexmengen $J \subseteq I$ gilt:

$$H_o^J = \bigcap_{j \in J} H_o^j \in \{H_o^i, i \in I\} \quad \text{oder} \quad H_o^J \text{ ist leer.}$$

Elementarhypothesen sind solche Hypothesen der Hypothesenfamilie, zu denen es keine echten Obermengen gibt, d.h. die keine andere Nullhypothese implizieren. In multiplen Tests bzgl. k Faktorstufen sind Elementarhypothesen oft Hypothesen bzgl. Paaren von Faktorstufen, in multiplen Tests bzgl. p abhängigen Variablen oft Hypothesen

bzgl. einzelner abhangiger Variablen; beim Vergleich gegen eine Kontrolle sind es alle Einzelvergleiche mit dieser Kontrolle.

Ein durchschnittsabgeschlossener multipler Test heißt konsonant, falls bei verworfenem H_o^j mindestens eine sie (echt) implizierende Hypothese H_o^i verworfen wird. Sonnemann (1982, Abschnitt 4) gibt ein einfaches Beispiel mit drei Elementarhypothesen.

Der bekannteste multiple Test zum multiplen Niveau α ist der sog. Bonferroni-Test:

Existiert für jedes von n Testproblemen mit Hypothesenfamilie $\{H_o^i, i=1,..,n\}$ ein Test ϕ_i zum Niveau α/n (d.h. $P_{H_o^i}(\phi_i=1) \leq \alpha/n$), ist $\phi=(\phi_1,..,\phi_n)$ ein multipler Test zum multiplen Niveau α.

Zum Beweis wird lediglich die Ungleichung von Boole bzw. Bonferroni (vgl. Hunter, 1976, S. 577) für die Wahrscheinlichkeit einer Vereinigung von Ereignissen benötigt. Deren Wahrscheinlichkeit ist kleiner oder gleich der Summe der Einzelwahrscheinlichkeiten. Damit hat man für jede Teilmenge $J \subseteq \{1,..n\}$:

Sei $H_o^J := \bigcap_{i \in J} H_o^i$ gültig; dann folgt:

$$(5.13) \quad P_{H_o^J}\left(\bigcup_{i \in J} (\phi_i=1)\right) \leq \sum_{i \in J} P_{H_o^i}(\phi_i=1) \leq |J| \cdot \alpha/n \leq \alpha \quad .$$

Zur Durchführung des Tests benötigt man nur die einzelnen Überschreitungswahrscheinlichkeiten (p-Werte) $p_i(x)$ nach (5.10).

Holm (1979) hat eine Modifikation des Bonferroni-Tests vorgeschlagen, die eine sequentielle Verwerfungsstrategie benutzt und eine gleichmäßige Verbesserung darstellt, d.h. mindestens alle mit dem Bonferroni-Test abgelehnten Teilhypothesen werden auch von diesem sog. Bonferroni-Holm-Test

abgelehnt. Für die p-Werte soll wie oben gelten:

(5.14) $P_{H_o^i}(p_i \leq \alpha/m) \leq \alpha/m$, $m \in \{1,..,n\}$ und alle $i=1,..,n$.

Zur Testdurchführung dieses multiplen Tests zum multiplen Niveau α werden die p_i aufsteigend nach ihrer Größe und mit ihnen die zugehörigen Hypothesen geordnet:

$$p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)} \quad .$$

Falls $p_{(1)} > \alpha/n$ (oder äquivalent $np_{(1)} > \alpha$) ist, werden alle Nullhypothesen $H_o^i, i=1,..,n$ akzeptiert und das Verfahren bricht ab. Ansonsten wird $H_1^{(1)}$ akzeptiert und im nächsten Schritt überprüft, ob $p_{(2)} > \alpha/(n-1)$ (bzw. $(n-1)p_{(2)} > \alpha$) ist. Falls ja, werden $H_o^{(2)},\ldots,H_o^{(n)}$ akzeptiert. Falls nicht, wird $H_1^{(2)}$ angenommen und $p_{(3)}$ mit $\alpha/(n-2)$ verglichen, usw. Eine Verbesserung gegenüber dem Bonferroni-Test ist aber nur möglich, wenn die erste 'Hürde' $p_{(1)} \leq \alpha/n$ überwunden wird. Die Abbildung 1 in Sonnemann (1981b) stellt das Verfahren in Form eines Flußdiagramms dar. Der Beweis, daß dieses sequentielle Verfahren ein multipler Test zum multiplen Niveau ist, ist leicht zu führen (vgl. Holm, 1979; Sonnemann, 1981b). Da die Beweisführung instruktiv für die folgenden Überlegungen ist, soll sie noch einmal dargestellt werden. Sei $J=\{i \in \{1,..,n\} \mid H_o^i \text{ ist wahr}\}$ und sei J o.B.d.A. nicht leer mit $m>1$ Elementen. Damit das multiple Niveau α eingehalten wird, muß man zeigen, daß folgende Bedingung erfüllt ist. Die Wahrscheinlichkeit, sich für alle H_o^i mit $i \in J$ zu entscheiden, muß großer oder gleich $1 - \alpha$ sein. Dazu muß man nachweisen, daß diese Wahrscheinlichkeit (1) größer oder gleich $P(\bigcap_{i \in J} \{p_i > \alpha/m\})$ und diese wiederum (2) großer oder gleich $1-\alpha$ ist.

Zu (1): Falls $p_i > \alpha/m$ für alle $i \in J$ Gültigkeit hat, muß auch $p_{(n-m+1)} > \alpha/m$ sein, denn von $p_{(n-m+1)}$ bis zum größten Wert $p_{(n)}$ sind es nur m-1 Elemente. Bis zum Schritt n-m+1 können nur Teilhypothesen mit $p_j \leq \alpha/(m-1)$ abgelehnt worden sein. Für diese tritt aber kein Fehler 1. Art auf, da in J nur Indizes sein sollen mit $p_i > \alpha/m$. Damit ist J eine Obermenge für alle $i \in J$ mit $p_i > \alpha/m$ und damit gilt (1).

Zu (2): Diese Abschätzung wird folgendermaßen geführt:

$$P\Big(\bigcap_{i\in J}\{p_i > \alpha/m\}\Big) = 1 - P\Big(\bigcup_{i\in J}\{p_i \leq \alpha/m\}\Big) \geq 1 - \sum_{i\in J} P(p_i \leq \alpha/m)$$

(5.15)

$$\overset{(5.14)}{\geq} 1 - |J| \cdot \alpha/m = 1 - m \cdot \alpha/m = 1 - \alpha \quad .$$

Falls zwei oder mehrere $p_{(i)}$ identisch sind, kann für sie eine beliebige Reihenfolge in der sequentiellen Prozedur gewählt werden.

Falls für einen Permutationstest ein Niveau α/m, $m \in \{1,..,n\}$ nicht erreichbar ist, weil die Menge zulässiger Permutationen weniger als m Elemente enthält, schlagen Petrondas & Gabriel (1983) vor, die betreffende(n) Nullhypothese(n) zu akzeptieren, da die Daten nicht genug Information zu ihrer Verwerfung enthalten.

Sonnemann (1981b) weist darauf hin, in praktischen Anwendungen der Prozedur als Hypothesen der Hypothesenfamilie Elementarhypothesen zu verwenden. Der Grund dafür wird in den folgenden Ausführungen erläutert.

Im Bonferroni-Holm Test werden nur (5.14) und die Boole'sche Ungleichung für die Wahrscheinlichkeit eine Vereinigung von Ereignissen verwendet. Eine Verbesserung der Abschätzung in (5.15) ist aber u.U. möglich, denn die Ungleichung

(5.16) $P(\bigcup_{i\in J}\{p_i\leq\alpha/|J|\})=P(\min_{i\in J} p_i\leq\alpha/|J|)\leq\sum_{i\in J}P(p_i\leq\alpha/|J|)=\alpha$

kann auf drei Arten verbessert werden:

- durch Verbesserung der Bonferroni-Ungleichung (Hunter, 1976; Worsley, 1982; Bauer & Hackl, 1985),
- durch Verkleinerung der $\alpha_i=\alpha/m$ für bestimmte Hypothesenfamilien (Hommel, 1984, 1985; Shaffer, 1984, 1986),
- durch Bestimmung der exakten gemeinsamen Verteilung der p_i und deren Minimum für alle $i\in J$.

Wenn es dabei möglich ist, kritische Werte $c_{\alpha,|J|}\geq\alpha/|J|$ mit den Eigenschaften:

(5.17) und
$$c_{\alpha,1}\geq c_{\alpha,2}\geq\ldots\geq c_{\alpha,n}$$
$$P_{\bigcap_{i\in J}H_o^i}(\min_{i\in J} p_i\leq c_{\alpha,|J|})\leq\alpha$$

zu finden, kann man das Bonferroni-Holm Verfahren mit den $c_{\alpha,m}$ anstelle von α/m benutzen. Auch das ergibt einen multiplen Test zum multiplen Niveau α (vgl. Sonnemann, 1981b).

Folgende Gesichtspunkte sind für dieses 'spezielle' Bonferroni-Holm Verfahren zu beachten:

(1) Falls $c_{\alpha,|J|}$ von der speziellen Zusammenstellung der Nullhypothesen H_o^i, $i\in J$, abhängt, wird

(5.18) $c_{\alpha,m}=\max\{c_{\alpha,m}(J),\ J\subset\{1,..,n\},\ |J|=m\}$

gewählt.

(2) Bei Permutationstests für Hypothesenfamilien bzgl. k Stufen eines Faktors (z.B.) in einem verteilungsfreien linearen Modell sind solche $c_{\alpha,m}$, die Informationen über die gemeinsame Verteilung mehrerer p_i, $i\in J$ nutzen, größer oder gleich $\alpha/|J|$ sind und (5.17) genügen, nicht

möglich. Wie in Petrondas & Gabriel (1983) dargestellt, werden z.B. für verschiedene Paare von Faktorstufen unabhängig voneinander nur die Beobachtungen unter den jeweiligen Paaren von Faktorstufen vertauscht.

(3) Für Hypothesenfamilien bezüglich der p Komponenten multivariater Beobachtungen kann obiges Verfahren in Permutationstests dagegen ideal benutzt werden, denn $\min_{i\in J} p_i$ ist exakt bestimmbar. Der erste Teil der Bedingung (5.17) muß jedoch in naheliegender Weise modifiziert werden. Darauf wird in Abschnitt 5.3 eingegangen.

Abschließend soll das Prinzip des allgemeinen Durchschnittstests erläutert werden (vgl. Sonnemann, 1982). Er wird in den Abschnitten 5.3 und 5.4 verwendet werden.

Nach (5.12) ist ein multipler Test $\phi=(\phi_1,..,\phi_n)$ genau dann ein Test zum multiplen Niveau α, wenn gilt:

$$P_{\bigcap_{i\in I} H_o^i}\left(\bigcup_{i\in I}\{\phi_i=1\}\right) \leq \alpha \quad .$$

Mit $I:=\{i\}$ sind alle ϕ_i Niveau-α-Tests für das jeweilige Testproblem: H_o^i vs. H_1^i. Aber auch für $\phi_I:=\max_{i\in I} \phi_i = \bigcup_{i\in I}\{\phi_i=1\}$ mit $I\subset\{1,..,n\}$ gilt:

Die ϕ_I sind Niveau-α Tests für $H_o^I:=\bigcap_{i\in I} H_o^i$ vs. $H_1^I:=\bigcup_{i\in I} H_1^i$.

Damit ist es sinnvoll, stets durchschnittsabgeschlossene Hypothesenfamilien anzunehmen, wobei für jede Nullhypothese H_o^i ein Niveau-α Test ϕ_i existieren soll; d.h.:

(5.19)

(1) $H_o^I:=\bigcap_{i\in I} H_o^i$ ist leer oder $H_o^I\in\{H_o^1,..,H_o^n\}$ für alle nichtleeren $I\subset\{1,..,n\}$

(2) $P_{H_o^i}(\phi_i=1) \leq \alpha$ mit $\phi_i=\begin{cases}1 & p_i \leq c_{i,\alpha}\\ 0 & p_i > c_{i,\alpha}\end{cases}$

Der allgemeine Durchschnittstest $\psi=(\psi_1,..., \psi_n)$, der nach Definition ein kohärenter Test zum multiplen Niveau α ist, wird aus $\phi=(\phi_1,..,\phi_n)$ folgendermaßen konstruiert (vgl. Einot

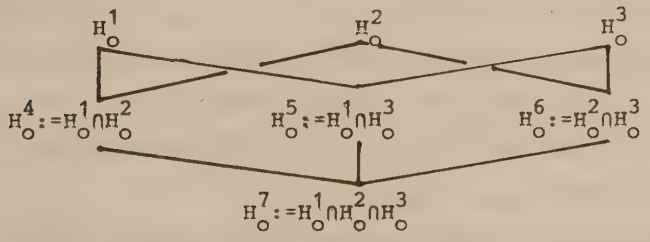
& Gabriel, 1975, Abschnitt 1.2; Petrondas & Gabriel, 1983, Abschnitt 3.1; Sonnemann, 1982, Abschnitt 4, Satz 4.3):

H_o^i wird genau dann verworfen, wenn $\tilde{\phi}_i$ und alle ϕ_j mit $H_o^j \subset H_o^i$ verworfen werden, d.h.

(5.20) $$\psi_i = \begin{cases} 1 & \text{für alle } H_o^j \subset H_o^i \text{ gilt: } \phi_j = 1, \text{ d.h. } \cap_{j: H_o^j \subset H_o^i} (\phi_j = 1) \\ 0 & \text{sonst} \end{cases}$$

Sonnemann (1982) weist darauf hin, daß die Komponenten des obigen allgemeinen Durchschnittstests üblicherweise nicht explizit bestimmt werden müssen. Bei endlichen, durchschnittsabgeschlossenen Hypothesenfamilien kann man sequentiell vorgehen.

Weiterhin sei vorausgesetzt, daß endlich viele Elementarhypothesen zur Erzeugung aller Hypothesen der Hypothesenfamilie vorhanden sind. Seien diese mit $H_o^1,..,H_o^m$, $2 \leq m \leq n$, bezeichnet. Beispielsweise sind für das m=p-variate unverbundene Zweistichprobenproblem die Hypothesen H_o^i: $\mu_{1i} = \mu_{2i}$ - i=1,..,p- elementar. Die durchschnittsabgeschlossene Hypothesenfamilie enthält $n-p=2^p-p-1$ aus diesen Hypothesen über Durchschnittsbildung erzeugte Hypothesen. Ordnet man die $n=2^p-1$ Hypothesen bezüglich der Mengeninklusion, ergibt sich für z.B. p=3 folgende Anordnung (ein sog. Hypothesenverband):



Für die Durchschnittshypothesen gilt:

$$H_o^i = \bigcap_{j=1}^{m} H_o^j \quad i=m+1,..,n \text{ mit } H_o^j \supseteq H_o^i \quad (\text{z.B. } H_o^4 = H_o^1 \cap H_o^2) \ ;$$

oder mit der Umindizierung $J(i)=\{j\in\{1,..,m\} \mid H_o^j \supseteq H_o^i\}$:

$$H_o^{J(i)} := \bigcap_{j\in J(i)} H_o^j = H_o^i \text{ für ein } i\in\{1,..,n\}$$

(z.B. $H_o^{J(1)} = H_o^1$, $H_o^{J(5)} = H_o^1 \cap H_o^3$) .

Tests für Durchschnittshypothesen wählt man folgendermaßen:

$$(5.21) \quad \phi_{J(i)} = \begin{cases} 1 & \leq \\ 0 & > \end{cases} \min_{j\in J(i)} p_j \quad c_{\alpha,J(i)}$$

Wenn man noch voraussetzt, daß die Vereinigungs-Durchschnittstests alle Tests zum Niveau α sind, d.h.

$$(5.22) \quad P_{H_o^{J(i)}}\left(\min_{j\in J(i)} p_j \leq c_{\alpha,J(i)} \right) \leq \alpha \quad ,$$

ist $\phi=(\phi_{J(1)},..,\phi_{J(n)})$ mit $J(n)\leq 2^m-1$ ein multipler Test zum multiplen Niveau α.

Aus ϕ läßt sich nach dem Abschlußprinzip in (5.20) wieder ein kohärenter Abschlußtest $\psi=(\psi_{J(1)},..,\psi_{J(n)})$ konstruieren (vgl. Sonnemann, 1981b, Abschnitt 5). Für die praktische Anwendung ist wichtig, daß der Abschlußtest nicht explizit angegeben werden muß, sondern sich aus der sequentiellen Durchführung der $\phi_{J(i)}$ ergibt.

Spezialfälle des allgemeinen Durchschnittstests sind der Bonferroni-Holm Test und spezielle sequentielle verwerfende Tests. Seien für die m Elementarhypothesen entweder mit dem Bonferroni-Holm Test oder einem speziellen Verfahren

folgende Entscheidungen getroffen worden (dabei sind die Hypothesen wie die p-Werte geordnet):

$$H_1^{(1)},..,H_1^{(m-k)},H_o^{(m-k+1)},..,H_o^{(m)} \quad .$$

Das multiple Niveau α bleibt eingehalten, wenn man zusätzlich die Durchschnittshypothesen $H_o^J = \cap_{i \in J} H_o^i$ mit $H_o^J \subset \cup_{j=1}^{m-k} H_o^{(j)}$ verwirft und alle übrigen Durchschnittshypothesen beibehält. Genauso kann man verfahren, wenn man irgendwelche Hypothesen und deren Durchschnitte betrachtet.

Der allgemeine Durchschnittstest und der sequentiell verwerfende Test gelangen bezüglich der Elementarhypothesen (bzw. Ausgangshypothesen) zum selben Ergebnis. Ein Beispiel gibt Sonnemann (1981b, Beispiel 1 in Abschnitt 2.3).

Für die Durchschnittstests nach (5.21) kann man im Fall des Bonferroni-Holm Tests $c_{\alpha,J(i)} = \alpha/|J(i)|$ wählen.

Komplette Hypothesenfamilien, d.h. Hypothesenfamilien, in denen alle aus Elementarhypothesen gebildeten Durchschnittshypothesen paarweise verschieden und nicht leer sind, lassen sich besonders leicht sequentiell testen. Beispiele für komplette Hypothesenfamilien sind die $n=2^p-1$ Hypothesen bezüglich Teilvektoren des p-dimensionalen Vektors in einem MANOVA-Testproblem oder auch die Vergleiche gegen eine Kontrollbedingung, in denen die Vergleiche jeder einzelnen Faktorstufe mit der Kontrolle die Elementarhypothesen darstellen. Für Permutationstests muß der letztere Fall jedoch ähnlich wie der Vergleich aller k Faktorstufen untereinander behandelt werden, da die zulässigen Permutationen unter der jeweiligen elementaren Nullhypothese stets nur die jeweilige einzelne Faktorstufe und die Kontrollbedingung betreffen.

Hypothesenfamilien mit paarweisen Vergleichen von Faktorstufen H_o^{ij}: $\mu_i=\mu_j$, $1\le i,j\le k$, sind nicht komplett, da nicht alle Schnitthypothesen verschieden sind. Für k=3 z.B. sind alle Durchschnitte von zwei H_o^{ij} identisch mit H_o^{123}:$\mu_1=\mu_2=\mu_3$. Wenn zusätzlich die Vereinigungs-Durchschnittstests konsonant sind, kann man folgendermaßen sequentiell testen:

Sei o.B.d.A. $p_1(x)\le..\le p_m(x)$, was durch Umnumerieren der m Elementarhypothesen immer zu erreichen ist:

(5.23)

(1) Gilt $p_1(x)> c_\alpha(\{1,..,m\})$, so ist $\psi_{J(i)}(x)=0$ für alle i=1,...,n und die sequentielle Prozedur wird gestoppt. Hat man dagegen $p_1(x)\le c_\alpha(\{1,..,m\})$, gilt $\psi_{J(i)}=1$ für alle J(i) mit $J(i)\cap\{1\}\neq\emptyset$.

Dann geht man über zu:

(2) Gilt $p_2(x)>c_\alpha(\{2,..,m\})$, so ist $\psi_{J(i)}=0$ für alle $J(i)\cap\{2,..,m\}$.
Falls $p_2(x)\le c_\alpha(\{2,..,m\})$ ist, hat man dagegen $\psi_{J(i)}=1$ für alle J(i) mit $J(i)\cap\{1,2\}\neq\emptyset$, usw.

(3) Sei l der Index, für den zum erstenmal tatsächlich $p_l(x)>c_\alpha(\{1,..,m\})$ ist. Dann werden alle $\psi_{J(i)}(x)=1$ mit $J(i)\cap\{1,..,l-1\}\neq\emptyset$ und alle $\psi_{J(i)}(x)=0$ mit $J(i)\cap\{1,..,m\}\neq\emptyset$ gesetzt.

Dieses Verfahren wird in abgewandelter Form in Abschnitt 5.3 Anwendung finden. Abschließend sei noch erwähnt, daß für komplette Hypothesenfamilien der obige Abschlußtest auch konsonant ist; i.a. muß das selbst bei konsonanten Vereinigungs-Durchschnittstests nicht der Fall sein (vgl. Beispiel 5.8 in Sonnemann, 1981b).

5.3 Multiple Permutationstests für die Komponenten multivariater Beobachtungen

5.3.1 Multiple sequentielle Permutationstests zum multiplen Niveau α

Sei irgendeines der MANOVA-Testprobleme aus Abschnitt 4.5 gegeben. Oft ist nicht nur von Interesse, ob es bezüglich der gesamten d-variaten Beobachtungsvektoren ein signifikantes Testergebnis gibt. Um im Folgenden Verwechslungen zu vermeiden, wird von d anstelle von p abhängigen Variablen ausgegangen. Vielmehr möchte man wissen, welche einzelnen oder Teilmengen der d abhängigen Variablen zur Verwerfung der entsprechenden Nullhypothesen führen. Man hat somit ein multiples Testproblem für die komplette Hypothesenfamilie aller 2^d-1 Hypothesen bezüglich Teilvektoren der d abhängigen Variablen. Auch wenn u.U. nur die Testergebnisse bezüglich der d Elementarhypothesen für einzelne Variablen eigentlich interessieren, wird sich zeigen, daß man die übrigen Hypothesen bezüglich 2- bis (d-1)-komponentiger Teilvektoren ähnlich wie in Abschnitt 5.2 mit hinzunehmen kann. Seien die $n=2^d-1$ Hypothesen so durchnumeriert, daß $H_o^1,\ldots,H_o^d$ die Elementarhypothesen bezüglich einzelner Komponenten bezeichnen und H_o^n die 'globale' Hypothese bezüglich aller d Komponenten. Da für die analogen univariaten Testprobleme exakte Permutationstests zum Niveau α existieren (vgl. Abschnitt 3.1.5), können der Bonferroni- und der Bonferroni-Holm Test stets angewendet werden, wenn α/m, $m=1,\ldots,d$, erreichbare Niveaus für die univariaten Permutationstests darstellen. Eine Verbesserung dieser beiden multiplen Tests zum multiplen Niveau α kann erzielt werden,

da das Minimum der p_i in (5.16) stets exakt bestimmbar ist. Dieser 'spezielle' sequentiell verwerfende Test soll im Folgenden dargestellt werden.

Sei $X_N=(X'_1,..,X'_N)'$ der Nd-dimensionale Vektor der d-dimensionalen Beobachtungen u.U. nach (maximal-)invarianter Transformation zur Beseitigung störender Effekte. Sei $X_N^i=(X_{1i},..,X_{Ni})'$, i=1,..,d, der Vektor der i-ten abhängigen Variablen. Sei weiterhin die Nd-dimensionale Verteilungsfunktion $F_N(x)$ von X_N mit $x\in R^{Nd}$ bei Gültigkeit der 'globalen' Nullhypothese H_o^n invariant unter allen Vektorpermutationen aus $\mathcal{Q}_N=Q_N\otimes\{1_d\}$ einer Untergruppe von $\mathfrak{S}_N$. Bei Gültigkeit der Elementarhypothesen $H_o^1,..,H_o^d$ gilt dasselbe für Q_N und jede der Komponenten $F_N^i(x)$ von F_N, i=1,..,d, $x\in R^N$.

Sei $S=(S_1,..,S_d)'$ der Vektor der o.B.d.A. identisch gewählten univariaten Teststatistiken. Diese haben dennoch unter der jeweiligen H_o^i nicht notwendig dieselbe (bedingte) Verteilung. Deshalb geht man von den S_i zu den zugehörigen Überschreitungswahrscheinlichkeiten über. Sei

$$p_i(y)=1/|Q_N|\cdot|\{\pi\in Q_N|S_i(\pi x_N^i)\geq y\}|,\ y\in R^1,\ i=1,..,d\ ,$$

die sog. (exakte) p-Funktion. Dann ist $p_i=p_i(S_i(x_N^i))$ der (univariate) p-Wert. Alternativ kann man die p-Funktion auch definieren als:

$$p_i(u) = P_{H_o^i}(p(S_i)\leq u)\quad ,\ u\in[0,1]\quad ;$$

Sie ist monoton fallend in y.

Sei $K=\{h_1,..,h_k\}$ eine nichtleere Teilmenge von $\{1,..,d\}$. Dann ist die (exakte) multiple p-Funktion (für Permutationstests) definiert als:

(5.24) $$p^{[K]}(u)=1/|\Omega_N|\cdot|\{\pi\in\Omega_N|\min_{h_\nu\in K}\{p_{h_\nu}(S_{h_\nu}(\pi x_N^{h_\nu}))\le u\}|\ ,\ u\in[0,1]$$

unabhängig davon, ob für Indizes j aus $\{1,..,d\}\setminus K$ H_o^j oder H_1^j gilt. Man hat damit:

$$p^{[K]}(u) = P_{\bigcap_{h_\nu\in K} H_o^{h_\nu}}(\min_{h_\nu\in K} p_{h_\nu}(S_{h_\nu})\le u)$$

$$=:P^K(\min_{h_\nu\in K} p_{h_\nu}(S_{h_\nu})\le u) \quad .$$

Da im Folgenden nur multiple sequentielle Testverfahren verwendet werden, ist nur folgender Spezialfall einer (exakten) multiplen p-Funktion von Interesse. Da die $P_i(S_i)$ unter der jeweiligen Nullhypothese H_o^i, $i=1,..,d$, alle die gleiche (Rechteck-) Verteilung auf $[0,1]$ besitzen, wird jeweils mit dem kleinsten univariaten exakten p-Wert begonnen.

Sei $p_1\le,.....,\le p_d$ die Ordnung der univariaten exakten p-Werte u.U. nach Umnumerierung der abhängigen Variablen. Dann definiert man:

$$p^{[i]}(u)=1/|\Omega_N|\cdot|\{\pi\in\Omega_N|\min_{i\le j\le d}\{p_j(S_j(\pi x_N^j))\}\le u\}|\ ,\ u\in[0,1]$$

(5.25) und als multiplen exakten p-Wert:

$$p^{[i]}=1/|\Omega_N|\cdot|\{\pi\in\Omega_N|\min_{i\le j\le d}\{p_j(S_j(\pi x_N^j))\}\le p_i\}|\ ,\ i=1,..,d\ .$$

Es werden also die Permutationen der abhängigen Variablen gezählt, die simultan in den Komponenten $i=1,..,d$ zu Werten der univariaten p-Funktionen führen, die kleiner oder gleich dem i-ten geordneten univariaten p-Wert sind (vgl. Maurer, 1981, Kap. 2.1; ohne Jahr).

Der neue sequentielle multiple Test zum multiplen Niveau α läßt sich dann in folgendem Flußdiagramm wie in Abbildung 5.1 darstellen:

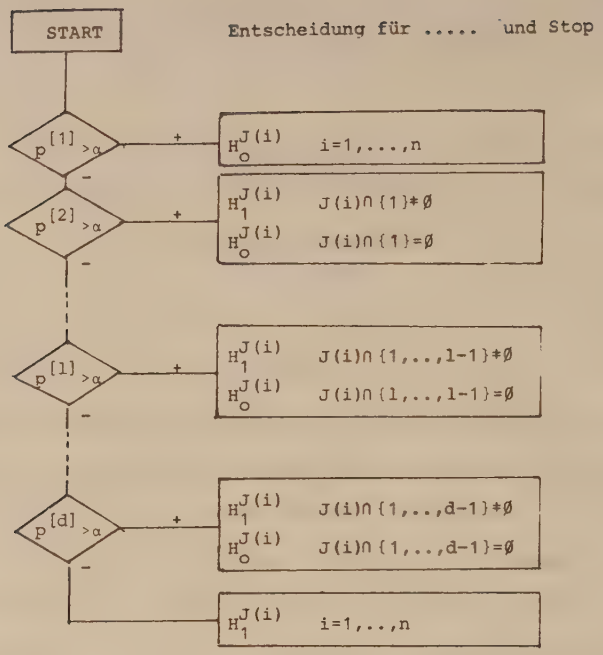


Abb. 5.1.: Sequentieller multipler Permutationstest zum multiplen Niveau α für die d Komponenten eines multivariaten Permutationstests mit Definition der exakten p-Werte nach (5.25); Numerierung der Elementar- und Schnitthypothesen nach Abschnitt 5.2.

Oft ist man nur an den Elementarhypothesen interessiert. Dann kann man in obiger Abbildung J(i)=1,..n in offensichtlicher Weise durch (Teil-)mengen der d Elementarhypothesen ersetzen. Vor dem Nachweis, daß diese sequentielle multiple Testprozedur das multiple Niveau α einhält, soll ein einfaches Beispiel für ihre Anwendung gegeben werden. Hartje & Willmes (1985) haben die Veränderungen der Leistungen von 10 aphasischen Patienten in d=6 Kontrolltests (LPS Untertests 3, 9, 10; BET Untertests 1, 2, 5) vor und nach einer Phase lateralisierter perzeptueller Stimulierung der rechten Hemi-

sphäre untersucht. Getestet werden sollte die einseitige Hypothese, daß eine Verbesserung der Leistungen von Vor- zu Nachtest eingetreten ist. Um die Anzahl zulässiger Permutationen in diesem d=6-variaten verbundenen Zweistichprobenproblem für die Beispielrechnung (!) nicht zu groß werden zu lassen, wurden per Zufall n=5 Patienten ausgewählt. Ihre Leistungen in Vor- und Nachtest sind in Tabelle 5.1 angegeben.

	VORTEST						NACHTEST					
	LPS 3	9	10	BET 1	2	5	LPS 3	9	10	BET 1	2	5
Patient	1	2	3	4	5	6	1	2	3	4	5	6
2 (1)	14	5	9	15	8	6	19	9	14	16	5	6
3 (2)	12	6	10	9	6	13	15	11	13	14	7	7
4 (3)	20	7	15	28	13	17	23	10	19	34	12	22
8 (4)	18	13	12	26	14	21	27	24	15	20	19	20
9 (5)	13	9	7	17	13	8	20	9	12	15	14	8
Median	14	7	10	17	13	13	20	10	14	16	12	8

Tab. 5.1.: Leistungen von 5 zufällig ausgewählten aphasischen Patienten in einer Therapiestudie zur lateralisierten Stimulierung der rechten Hirnhälfte (Hartje & Willmes, 1985); Kontrolltests sind je drei Untertests aus dem LPS und dem BET; angegeben sind Rohwerte.

Als Teststatistik für die univariaten einseitigen verbundenen Zweistichprobenprobleme eignet sich

$$S_i(X) = \sum_{j=1}^{5} X_{2ji} \quad , \; i=1,..,6; \; j=1,..,5 \; \text{(vgl.(3.39))} \; .$$

In Tabelle 5.2 ist der Vektor der Teststatistikwerte für jede der $2^5=32$ unter der Nullhypothese zulässigen Vertauschungen $\mathfrak{Q}_n = Q_n \otimes \{I_6\}$ der Rohwertvektoren aus Vor- und Nachtest angegeben. Je Zeile zeigt der 5-stellige Indikatorvektor in Tab. 5.2 an, ob Vor- und Nachtestleistung eines Patienten vertauscht (=1) sind oder nicht (=0). Der Indikatorvektor

(0,0,0,0,0) kennzeichnet den Teststatistikvektor fur die unpermutierten Beobachtungen. Die Werte fur die Permutationen l=17,..,32 ergeben sich aus denen fur die Permutationen l=1,..16, wenn man die Werte für den komplementären Indikatorvektor (d.h. 0 und 1 sind vertauscht) von der Summe $\sum_{j=1}^{5}(X_{1ji}+X_{2ji})$ subtrahiert.

Permutations Index	Permutat.-Indik. j=1	2	3	4	5	Werte der univariaten Teststatistiken S_i i= 1	2	3	4	5	6
l= 1	0	0	0	0	0	104	63	73	99	57	63
2	1	0	0	0	0	99	59	68	98	60	63
3	0	1	0	0	0	101	58	70	94	56	69
4	0	0	1	0	0	101	60	69	93	58	58
5	0	0	0	1	0	95	52	70	105	52	64
6	0	0	0	0	1	97	63	68	101	56	63
7	1	1	0	0	0	96	54	65	93	59	69
8	1	0	1	0	0	96	56	64	92	61	58
9	1	0	0	1	0	90	48	65	104	55	64
10	1	0	0	0	1	92	59	63	100	59	63
11	0	1	1	0	0	98	55	66	88	57	64
12	0	1	0	1	0	92	47	67	100	51	70
13	0	1	0	0	1	94	58	65	96	55	69
14	0	0	1	1	0	92	49	66	99	53	59
15	0	0	1	0	1	94	60	64	95	57	58
16	0	0	0	1	1	88	52	65	107	51	64
17	1	1	1	0	0	93	51	61	87	60	64
18	1	1	0	1	0	87	43	62	99	54	70
19	1	1	0	0	1	89	54	60	95	58	69
20	0	1	1	1	0	89	54	63	94	52	65
21	0	1	1	0	1	91	55	61	90	56	64
22	0	0	1	1	1	85	49	61	101	52	59
23	0	1	0	1	1	85	47	62	102	50	70
24	1	0	1	1	0	87	55	61	98	56	59
25	1	0	0	1	1	83	48	60	106	54	65
26	1	0	1	0	1	89	56	59	94	60	58
27	1	1	1	1	0	84	40	58	93	55	65
28	1	1	1	0	1	86	51	56	89	59	64
29	1	1	0	1	1	80	43	57	101	53	70
30	1	0	1	1	1	80	45	56	100	55	59
31	0	1	1	1	1	82	44	58	96	51	65
32	1	1	1	1	1	77	40	53	95	84	65
	$\sum_{j=1}^{5}(x_{1ji}+x_{2ji})$					181	103	126	194	111	128

Tab. 5.2.: Univariate Teststatistikwerte für die Beispieldaten aus Hartje & Willmes (1985) unter allen zulässigen $2^5=32$ Permutationen im verbundenen Zweistichprobenproblem; Teststatistik ist $S_i(X)=\sum_j X_{2ji}$, i=1,..,6, die Summe der Nachtestwerte.

Ordnet man je Dimension die Teststatistikwerte nach aufsteigender Größe wie in Tabelle 5.3, kann man die univariaten p-

Werte p_i leicht ablesen. Die geordneten univariaten p-Werte sind:

$p_{(1)}=p_1=1/32; p_{(2)}=p_3=1/32; p_{(3)}=p_2=2/32$

$p_{(4)}=p_5=12/32; p_{(5)}=p_4=14/32; p_{(6)}=p_6=24/32.$

Die p-Werte p_1 und p_3 hätten auch vertauscht werden können, ohne daß sich an den Resultaten der multiplen Testprozedur etwas ändert (vgl. die Argumentation in Sonnemann, 1981b, Abschnitt 2.1, Bemerkung 1). Die multiplen p-Werte kann man sequentiell leicht mit den Tabellen 5.2 und 5.3 ermitteln.

	Abhängige Variable							
Index	1	2	3	4	5	6		
1	77	40	53	87	50	58		
2	80	40	56	88	51	58		
3	80	43	56	89	51	58		
4	82	43	57	90	51	58		
5	83	44	58	92	52	59		
6	84	45	58	93	52	59		
7	85	47	59	93	52	59		
8	85	47	60	93	53	59	p_6	(24/32)
9	86	48	60	94	53	63		
10	87	48	61	94	54	63		
11	87	49	61	94	54	63		
12	88	49	61	95	54	63		
13	89	51	61	95	55	64		
14	89	51	62	95	55	64		
15	89	52	62	96	55	64		
16	90	52	63	96	55	64		
17	91	54	63	98	56	64		
18	92	54	64	98	56	64	p_4	(14/32)
19	92	54	64	99	56	64		
20	92	55	65	99	56	64	p_5	(12/32)
21	93	55	65	99	57	65		
22	94	55	65	100	57	65		
23	94	56	65	100	57	65		
24	95	56	66	100	58	65		
25	96	58	66	101	58	69		
26	96	58	67	101	59	69		
27	97	59	68	101	59	69		
28	98	59	68	102	59	69		
29	99	60	69	104	60	70		
30	101	60	70	105	60	70	p_2	(2/32)
31	101	63	70	106	60	70	p_1, p_3	(1/32)
32	104	63	73	107	61	70		

Tab. 5.3.: Geordnete univariate Teststatistikwerte für die Beispieldaten zusammen mit den univariaten p-Werten.

i=1: Nach (5.25) muß man zur Bestimmung von

$$p^{[1]}=1/32\,|\{\pi\in Q_n \mid \min_{1\le j\le 6}\{p_j(S_j(\pi x_n^j))\}\le p_{(1)}\}|$$

alle Mengen $M_j^1, j=1,..,6$ von Permutationen π_1, 1=1,..,32, ermitteln, die zu Teststatistikwerten mit zugehörigen p-Werten kleiner oder gleich $p_{(1)}=1/32$ führen. Sicher gehört $M_1^1=\{\pi_1\}$ dazu, denn der univariate p-Wert $p_{(1)}$ für die abhängige Variable ergibt sich gerade für diese Permutation (vgl. Tab. 5.3 erste Spalte). Seien im Folgenden zur Vereinfachung der Notation nur die Indizes der Permutationen in den Mengen angegeben. Bei Variable 3 ist es ebenfalls π_1, d.h. $M_3^1=\{1\}$, bei Variable 4 $M_4^1=\{16\}$ und bei Variable 5 $M_5^1=\{8\}$. Für die abhängigen Variablen 2 und 6 sind dagegen die minimalen erreichbaren p-Werte 2/32 bzw. 4/32, sodaß bei der Bildung des Minimums keine weiteren Permutationen berücksichtigt werden müssen, d.h. $M_2^1=\emptyset$, $M_6^1=\emptyset$.

Insgesamt hat man als zu berücksichtigende Permutationen

$$M^{[1]}=M_1^1\cup M_3^1\cup M_2^1\cup M_5^1\cup M_4^1\cup M_6^1=\{1,8,16\}$$

und damit $p^{[1]}=3/32$.

Es sei hier darauf hingewiesen, daß der multiple p-Wert der Bonferroni-Holm Prozedur (vgl. Maurer, 1981, Lemma 3) echt größer ist: $p_{BH}^{[1]}=6\cdot p_{(1)}=6/32$ (vgl. Abschnitt 5.2).

i=2: Nur die abhängigen Variablen 2-6 müssen noch berücksichtigt werden. Da $p_{(1)}=p_{(2)}$ ist, hat man $M^{[2]}=M^{[1]}$ und $p^{[2]}=3/32<p_{BH}{}^{[2]}=5/32$.

i=3: Es werden die Variablen 2, 4, 5, 6 berücksichtigt. Für Variable 2 sind die Permutationen π_1 und π_6 zu notieren, für Variable 5 bleibt es lediglich π_8, da das zweitkleinste erreichbare Niveau für die 5. Variable $p_5=4/32$ ist. Für Variable 4 kommt π_{25} hinzu. M_6^3 bleibt weiterhin

leer, da das kleinste erreichbare Niveau 4/32 ist. Insgesamt ergibt das

$$M^{[3]} = M_2^3 \cup M_5^3 \cup M_4^3 \cup M_6^3 = \{1,6\} \cup \{8\} \cup \{16,25\} \cup \emptyset = \{1,6,8,16,25\}$$

Damit ist

$$p^{[3]} = 5/32 < p_{BH}^{[3]} = 4 \cdot p_{(3)} = 8/32 \quad .$$

Aufgrund des sequentiellen Vorgehens nach der Ordnung der univariaten p-Werte gilt für die Mengen von Permutationen je abhängiger Variabler stets:

$$M_j^l \subseteq M_j^{l+1} \quad , \; l=1,..,d-1 \; , \text{ für alle } j=1,..,d.$$

i=4: $M^{[4]} = M_5^4 \cup M_4^4 \cup M_6^4$

$= \{8,2,17,26,7,10,28,4,19,1,11,15\} \cup \{16,25,5,9,23,6,$
$22,29,10,12,30\} \cup \{12,18,23,29,3,7,13,19,20,27,31,32\}$

und damit

$$p^{[4]} = 29/32 \quad .$$

i=5: $M^{[5]} = M_4^5 \cup M_6^5 = M_4^4 \cup \{1,14,18\} \cup M_6^5$

$$p^{[5]} = 22/32 \quad .$$

i=6: $p^{[6]} = p_{(6)} = 24/32 \quad .$

Für ein multiples Niveau $\alpha = 5/32$ wären somit nach dem Bonferroni-Holm Test keine, nach diesem neuen, von Maurer (1981, S. 30) angedeuteten sequentiellen Verfahren aber die Leistungen in den abhängigen Variablen 1-3 als signifikant verbessert anzusehen. Berücksichtigt man alle 10 aphasischen Patienten, so sind die exakten multiplen p-Werte in Tabelle 5.4 zusammen mit den multiplen p-Werten des Bonferroni-Holm Tests angegeben. Sie wurden nach dem in Abschnitt 8.4.4 beschriebenen Algorithmus bestimmt.

Abhängige Variable	Exakte* multiple p-Werte $p^{[i]}_{i=1,..,6}$	Multiple p-Werte nach Bonferroni-Holm
1	.00293	.00586
3	.00293	.00488
2	.00488	.00781
5	.07422	.09082
4	.05371	.06250
6	.06250	.06250

* 2^{10}=1024 zulässige Permutationen

Tab. 5.4.: Ergebnisse des exakten multiplen Permutationstests zum multiplen Niveau α für ein 6-variates verbundenes Zweistichprobenproblem (einseitiger Test, Daten aus Hartje & Willmes (1985))

Man gelangt zwar für ein multiples α=5% mit beiden Verfahren zu denselben Testentscheidungen, die gleichmäßige Überlegenheit der exakten multiplen p-Werte gegenuber den multiplen p-Werten des Bonferroni-Holm Tests ist aber deutlich zu sehen.

Zum Beweis, daß es sich bei dem Verfahren aus Abbildung 5.1 um einen multiplen Test zum multiplen Niveau α handelt, überlegt man Folgendes:

(1) Die in Abschnitt 5.2 für die Durchführung der Vereinigungs-Durchschnittstests benötigten Schwellenwerte $c_{\alpha,J(i)}$ in Formel (5.21) bzw. die $c_{\alpha}(\{i,..,m\})$, i=1,..,m, für die komplette Hypothesenfamilie und konsonante Vereinigungs-Durchschnittstests aus (5.23) sind für Permutationstests prinzipiell nicht angebbar. Denn die gemeinsame Verteilung der $p_j, j \in J(i)$, und damit vom Minimum dieser p_j unter der jeweiligen Nullhypothese ist apriori nicht bekannt, da Permutationstests bedingte Tests sind. Die gemeinsame Verteilung ist nur bezuglich der beobachteten Kovariation der d abhängigen Variablen bestimmbar. Der Bonferroni-Holm Test mit der Wahl $c_{\alpha,J(i)}=\alpha/|J(i)|$ ist dagegen ohne diese Kenntnis

durchführbar (vgl. (5.23) und Sonnemann, 1982, Bemerkung 5.12).

(2) Der Beweis wird mit Hilfe der von Maurer (1981) vorgeschlagenen sog. exakten multiplen p-Werte gefuhrt (vgl. (5.25)). Seien die Elementarhypothesen $H_o^1,\ldots,H_o^d$ o.B.d.A. wie die univariaten p-Werte $p_{(1)}\leq\ldots\leq p_{(d)}$ geordnet. Da Permutationstests exakte Tests zum Niveau α sind, gilt bekanntlich:

$$(*)\quad P_{H_o^i}(p_{(i)}\leq\alpha)\leq\alpha\ ,\ i=1,..,d\quad .$$

Den Ausdruck (5.25) kann man auch schreiben als:

$$(**)\quad \underline{p}^{[i]} = P_{\bigcap_{j\in\{i,..,d\}} H_o^j}(\min_{i\leq j\leq d}\{p_j(S_j)\}\leq p_{(i)})$$
$$=: P^i(\min_{i\leq j\leq d}\{p_j\}\leq p_{(i)})\quad ,\ i=1,..,d.$$

unabhängig davon, ob H_o^j oder H_1^j gilt für $j\notin\{i,..,d\}$.

(3) Es gilt (vgl. Maurer, 1981, Lemma 1):

$$P_{\bigcap_{j\in\{i,..,d\}} H_o^j}(\underline{p}^{[i]}\leq\alpha)\ \leq\ P_{\bigcap_{j\in\{i,..,d\}} H_o^j}(p_{(i)}\leq\alpha)\ \leq\ \alpha,\alpha\in(0,1),$$

aufgrund der Definition der multiplen p-Werte und der Tatsache, daß die Tests für die Elementarhypothesen Niveau-α Tests sind.

(4) Für $\emptyset\neq K:=\{h_1,..,h_k\}\subset\{i,..,d\}$, $k<d-i+1$, $i=1,..,d$,
gilt $\underline{p}^{[K]}\leq\underline{p}^{[i]}$.

Per Definition ist:

$$\underline{p}^{[K]}(u) = P^K(\min_{h_\nu\in K}\{p_{h_\nu}(S_{h_\nu})\}\leq u)$$
$$\overset{(2)}{=} P^i(\min_{h_\nu\in K}\{p_{h_\nu}(S_{h_\nu})\}\leq u)$$
$$\leq P^i(\min_{j\in\{i,..,d\}}\{p_j(S_j)\}\leq u)\ ,\ \text{da } K\subset\{i,..,d\} \text{ und } p_{(i)}\leq\ldots\leq p_{(d)}\ .$$
$$= \underline{p}^{[i]}(u)$$

Damit ist die Konsonanz der Testentscheidung nachgewiesen. Falls $K \subset J$, d.h. $H_o^{[J]} := \bigcap_{j \in J} H_o^j \subset \bigcap_{j \in K} H_o^j =: H_o^{[K]}$, impliziert die Ablehnung von mindestens einem $H_o^{[J]}$ mit $K \subset J$, d.h. $p^{[J]} \leq \alpha$, die Ablehnung von $H_o^{[K]}$, da $p^{[K]} \leq p^{[J]}$.

(5) Kohärenz der Testentscheidungen ist durch das sequentielle Vorgehen sichergestellt.

(6) Den Nachweis, daß das Testverfahren für i=1,..,d das multiple Niveau α einhält, führt man wie bei Maurer (1981, Satz 1).

(7) Die multiplen p-Werte erfüllen alle Voraussetzungen des Satzes 5.10 in Sonnemann (1982), wenn man dort die Formel (5.5) durch die Vereinigungs-Durchschnittstests

$$\phi_{J(i)} = \begin{cases} 1 & p^{[J(i)]} \leq \\ & \\ 0 & p^{[J(i)]} > \end{cases} \alpha$$

ersetzt. Diese $\phi_{J(i)}$ sind nach (3) und (4) Niveau-α Tests, d.h. $P^{J(i)}(p^{[J(i)]} \leq \alpha) \leq \alpha$.

Wegen (4) gilt auch: $H_o^{J(i)} \subset H_o^{J(j)} \Rightarrow p^{[J(j)]} \leq p^{[J(i)]}$,

sodaß man den kohärenten und konsonanten Abschlußtest zum multiplen Niveau α in Analogie zur Formel (5.7) in Sonnemann (1982) konstruiert:

$$\{\psi_{J(i)}=1\} = \bigcap_{J(j) \supseteq J(i)} \{\phi_{J(j)}=1\} = \bigcap_{J(j) \supseteq J(i)} \{p^{[J(j)]} \leq \alpha\}$$

(8) Die Bemerkung 5.11 bei Sonnemann (1982) sichert, daß die Prozedur aus Abbildung 5.1 korrekt ist, wenn man die Abfragen bei Sonnemann $p_i \overset{>}{\underset{<}{=}} c_\alpha(\{i,..,m\})$ durch $p^{[i]} \overset{>}{\underset{<}{=}} \alpha$ ersetzt.

Aufgrund der Schritte (7) und (8) im obigen Beweis kann man in dem dargestellten Beispiel unter Kontrolle des multiplen Niveaus α nicht nur die Elementarhypothesen H_o^1, H_o^2, H_o^3 verwerfen sondern auch alle Durchschnittshypothesen, die mindestens einen der Indizes 1, 2, 3 enthalten.

Mit obigem Verfahren aus Abbildung 5.1 ist ein allgemeines Konstruktionsprinzip für multiple Permutationstests bezüglich der Komponenten multivariater Beobachtungen gegeben. Es kann für alle multivariaten linearen Modelle aus Abschnitt 4.5 angewendet werden, indem man zum Testen der Elementarhypothesen die entsprechenden univariaten Permutationstests aus Abschnitt 3.1.5 verwendet.

Es sei nur erwähnt, daß man wie bei Maurer (1981) anstelle der exakten multiplen p-Werte in (5.25) die folgenden multiplen p-Werte hätte benutzen können:

Sei $\tilde{p}_{gh,i} := 1/|\Omega_N| \cdot |\{\pi \in \Omega_N | p_g(S_g(\pi x_N^g)) \leq p_{(i)} \wedge p_h(S_h(\pi x_N^h)) \leq p_{(i)}\}|$

und

$$\tilde{p}^{[i]} := (d-i+1)p_{(i)} - \sum_{(g,h) \in B_{max}^{[i]}} \tilde{p}_{gh,i} \quad , i=1,..,d,$$

wobei $B_{max}^{[i]}$ der maximale Pfad in einem Graphen mit der Knotenmenge $\{i,..,d\}, i=1,..,d,$ und den Kantenlängen $\tilde{p}_{gh,i}$ - $g,h \in \{i,..,d\}$ - bezeichne (vgl. auch Hunter, 1976; Worsley, 1982; Bauer & Hackl, 1985).

In diesem Verfahren wird eine Verbesserung der multiplen p-Werte aus dem Bonferroni-Holm Verfahren $p_{BH}^{[i]} = (d-i+1)p_{(i)}$ erzielt durch Nutzung der Kovariation aller Paare von Variablen. Im exakten Permutationstest bedeutet es aber keinen großeren Aufwand, sofort die Kovariation aller verbliebenen Variablen über $\min_{i \leq j \leq d} p_j$,i=1,..,d, auszunutzen. Häufig dürfte der Unterschied zwischen $\tilde{p}^{[i]}$ und dem exakten $p^{[i]}$ nicht sehr groß sein. Gegenüber den multiplen p-Werten nach dem Holm'schen Verfahren bedeuten die exakten multiplen p-Werte in Fällen hoher Korrelation zwischen den abhängigen Variablen besonders in den ersten Schritten der sequentiell verwerfenden Prozedur eine große Verbesserung der Teststärke, da dann $p^{[i]} \approx p_{(i)}$. Für alle abhängigen Variablen

erfullen dann annahernd dieselben Permutationen die Bedingung (5.25). Somit sind kaum mehr Permutationen als bei $p_{(i)}$ allein zu berücksichtigen. Für den Fall perfekter Korrelation (positiv oder negativ für zweiseitige Tests) gilt sogar exakt $p^{[i]}=p_{(i)}$ für alle $i=1,..,d$, sodaß die $p_{(i)}$ stets mit α und nicht mit $\alpha/(d-i+1)$ zu vergleichen sind. Stoline & Mitchell (1981) haben gezeigt, daß für hohe Korrelationen im parametrischen multivariaten Zweistichprobenproblem das exakte Verfahren der Hunter'schen Verbesserung der Bonferroni-Ungleichung deutlich überlegen ist.

Auch für die MRPP-Statistiken aus Abschnitt 4.7 läßt sich das Verfahren direkt benutzen, wenn man den exakten Permutationstest und nicht die approximative Losung mit Hilfe einer Pearson-III Verteilung wählt. Dazu betrachtet man z.B. für das unverbundene k Stichprobenproblem den Vektor der univariaten Statistiken

(5.27) $\delta=(\delta_1(Y^1),..,\delta_d(Y^d))$ mit

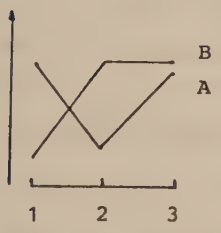
$$\delta_l(Y^l)=\sum_{i=1}^{k}(n_i/n)\cdot(n_i(n_i-1)/2)\sum_{j<j'}^{n_i}|Y^l_{ij}-Y^l_{ij'}|^v \quad \text{mit } v=1 \text{ oder } 2,\ l=1,..,d;\ n=\sum_i n_i$$

und

$$p_l(x)=1/|S_n|\cdot|\{\pi\in S_n|\delta_l(\pi y^l)\le x\}| \quad , \ x\in[0,\infty) \ , \ l=1,..,d \ .$$

Für (M)RPP-Statistiken sprechen kleine Teststatistikwerte gegen die Nullhypothese der Vertauschbarkeit der Beobachtungen zwischen den Faktorstufen. Die multiplen p-Werte bleiben wie in (5.25) definiert mit δ_l anstelle von S_i. Ob es auch möglich ist, approximative multiple p-Werte $\hat{p}^{[i]}=P(\delta^{[i]}\le\delta_{(i)})$ für abhängige approximativ Pearson-III-verteilte Zufallsvariablen zu bestimmen, soll hier nicht untersucht werden.

In der multivariaten Analyse von Profilen sind obige multiple Tests zum multiplen Niveau α nicht direkt einsetzbar. Problematisch erscheint z.B. die Angabe einer geeigneten Hypothesenfamilie. Beim multivariaten Vergleich von k Profilen mit den d-1 Differenzen benachbarter Profilstufen als Kontraste ist z.B. von Interesse, welche Profilstufen-Differenzen für die global festgestellte fehlende Parallelität der Profile 'verantwortlich' sind. Für den Fall zweier Plotstufen A, B mit d=3 Profilelementen, d.h. zwei Differenzen: Stufe 1 - Stufe 2, Stufe 2 - Stufe 3, erhält man je nach Wahl eines ein- oder zweiseitigen (Permutations-) Tests für die entsprechenden ein- oder zweiseitigen Hypothesen unterschiedliche Ergebnisse, wie man sich anhand folgender Skizze leicht überlegt:



H_o^{12}: $\mu_{A1}-\mu_{A2} \lesseqgtr \mu_{B1}-\mu_{B2}$ verwerfen, aber

$|\mu_{A1}-\mu_{A2}| \neq |\mu_{B1}-\mu_{B2}|$ akzeptieren

Schon für den Fall des k Stichprobenproblems mit einseitigen Hypothesen ist bislang keine zufriedenstellende Lösung gefunden worden (vgl. Bemerkung 6.2 in Sonnemann, 1982), auch wenn Bauer et al. (1986) ein sehr konservatives Verfahren für Paare von einseitigen Hypothesen vorschlagen.

3.2 Simultane Testprozeduren für Permutationstests

Wie in Abschnitt 5.1.4 erläutert, sind mit Permutationstests nur Hypothesen bezuglich der d Komponenten der abhängigen Variablen sinnvoll testbar, sofern es sich um simultane Testprozeduren handelt. Man muß sich jedoch nicht auf die Familie aller 2^d-1 Hypothesen bezüglich der Komponenten der Beobachtungsvektoren beschränken, sondern kann auch Hypothesen über Linearkombinationen oder (orthogonale) lineare Kontraste zwischen den abhängigen Variablen betrachten (als Spezialfall von $\mathfrak{H}$ mit $G=G_o$ in Gabriel, 1969b, S. 70).

Unter der globalen Nullhypothese sowie unter jeder Teilhypothese sind dieselben Permutationsgruppen - u.U. nur bezüglich eines Teilvektors aller abhängigen Variablen - zulässig, da für alle dieselben Faktorstufen(kombinationen) untereinander verglichen werden.

Zur Darstellung der Verfahren für Permutationstests genügt es, das unverbundene d-variate k-Stichprobenproblem zu betrachten. In mehrfaktoriellen Versuchsplänen muß u.U. eine (maximal-) invariante Transformation des gesamten Beobachtungsvektors vorgeschaltet werden, um für die jeweils betrachtete Hypothesenfamilie 'störende' Parameter zu eliminieren.

Sei die globale Nullhypothese in einem verteilungsfreien Lokationsmodell

$$H_o: \ \mu_i=\mu_G \quad \text{für alle } i\in G\subseteq\{1,..,k\},\ |G|=g\geq 2,\ \mu_i\in R^d \ .$$

Neben den Hypothesen bezüglich der einzelnen Komponenten und deren Durchschnitten seien aber auch Hypothesen der Form:

$$H_o^a: \quad a'\mu_i = a'\mu_G \quad , \; a \in R^d, \; a \neq 0 \; ; \; H_o = \bigcap_{a \neq 0} H_o^a$$

(5.28) oder

$$H_o^A: \quad A'\mu_i = A'\mu_G \quad \text{mit der Kontrastmatrix A ,}$$

$$A \in \{C_{d \times d-1} \mid 1_d'C = 0_{d-1}', \; 2 \leq rg(C) \leq d-1\}$$

von Interesse.

Ein ausführlich besprochenes Beispiel mit G=6, d=3, $a=(1,1,1)'$ und

$A = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix}$ findet man bei Gabriel (1968, Abschnitt 2).

Wie in Abschnitt 5.1.3 für die Hypothesenfamilie Ω (vgl. Gabriel, 1969b) beschrieben, sind als Teststatistiken Funktionen der Eigenwerte von

$$\psi^a(Y) = \psi(((\underline{S}_e + \underline{S}_h)(a'Y))^{-1} \underline{S}_h(a'Y))$$

(5.29) bzw.

$$\psi^A(Y) = \psi((A'(\underline{S}_e + \underline{S}_h)(Y)A)^{-1} A'\underline{S}_h(Y)A)$$

geeignet (vgl. Gabriel, 1969b). Falls A nicht den Rang d-1 besitzt, kann die Moore-Penrose Inverse in der letzten Gleichung gewählt werden. Für Linearkombinationen a'Y sind $\underline{S}_e$ und $\underline{S}_h$ die Summen der quadrierten Abweichungen aus der ANOVA.

In einer simultanen Testprozedur werden die Prüfgrößen mit einem einzigen gemeinsamen Schwellenwert verglichen. Wie in Abschnitt 5.1.3 schon dargestellt, wird das entsprechende $(1-\alpha)$-Quantil der exakten Permutationsverteilung der Teststatistik $\psi(Y)$ als dieser Schwellenwert gewählt. Es werden alle Permutationen ausgeführt, die bei Gültigkeit der globalen Nullhypothese zulässig sind. Im Fall des unverbundenen multivariaten k Stichprobenproblems sind das alle Vertauchungen der d-dimensionalen Beobachtungen, von denen wie üblich nur $n_.!/\prod_i n_i!$ zu verschiedenen Teststatistikwerten führen können. Je nach Wahl von ψ kann es auch das

untere α-Quantil der exakten Verteilung sein. Falls 1 - α bzw. α kein erzielbares Niveau in der exakten Verteilung ist, wird das nächstgroßere Quantil zu 1-α, $\psi^{G}_{d,1-\alpha}$, bzw. das nächstkleinere zu α, $\psi^{G}_{d,\alpha}$ gewählt. Für alle Teilhypothesen bzgl. einer Teilmenge aller d abhängigen Variablen wird $\psi^{V}(Y)$, $V \subset \{1,..,d\}$ und für Teilhypothesen bzgl. Linearkombinationen oder (orthogonalen) Kontrasten in den Variablen wird $\psi^{a}(Y)$ bzw. $\psi^{A}(Y)$ für Y=y berechnet und mit dem entsprechenden (1 - α)-Quantil $\psi^{G}_{d,1-\alpha}$ bzw. $\psi^{G}_{d,\alpha}$ verglichen. Die Teilhypothese wird verworfen, falls $\psi^{V}(y)$ bzw. $\psi^{a}(y)$ bzw. $\psi^{A}(y)$ größer oder gleich $\psi^{G}_{d,1-\alpha}$ bzw. kleiner gleich $\psi^{G}_{d,\alpha}$ ist.

Nach Eigenschaft II in Gabriel (1968) halten die zu den Hypothesenfamilien gehörenden Tests das multiple α -Niveau ein, denn die Wahrscheinlichkeit mindestens einen Fehler 1. Art zu machen, ist genau α, wenn H_o gilt und sonst immer nicht größer als α. Für den Fall multivariat normalverteilter Beobachtungen ist bekannt, welche Funktionen der Eigenwerte für welche Hypothesenfamilie besonders 'gunstige' Eigenschaften besitzen. Kohärenz gilt für alle ψ (nach Eigenschaft I), die monoton wachsende Funktionen der Eigenwerte (sog. 'increasing root functions') sind (Gabriel, 1968, S. 491; 1969b, S. 76). Alle Teststatistiken aus (4.18) erfüllen diese Bedingung. Beide Maximalwurzelkriterien führen sogar zu einer konsonanten STP (Eigenschaft III). Für beliebige ψ^{a} oder ψ^{A} haben die Maximalwurzelkriterien das größte 'Auflösungsvermögen', d.h. sie verwerfen die meisten Elementarhypothesen (Eigenschaft V). Bezüglich der Teststärke sowohl für die globale Hypothese wie z.B. für die Elementarhypothesen gibt es keine eindeutigen Ergebnisse.

Falls Teststärke für die Elementarhypothesen wichtig ist, sind die Maximalwurzelkriterien wiederum besonders geeignet. Ob alle diese Eigenschaften auch für die STPen mit exakten Permutationstests gelten, ist nicht untersucht. Die Kohärenz- und Konsonanzaussagen sind übertragbar, denn in den Beweis (Gabriel, 1968, Appendix A, B) gehen lediglich algebraische Argumente ein.

In praktischen Anwendungen ist aber eine Aussage für alle ψ^a oder ψ^A u.U. nicht von Interesse. Für eine 'kleinere' Anzahl von solchen Hypothesen liefert die Anwendung des Bonferroni-Holm Tests mit großer Wahrscheinlichkeit 'günstigere' Ergebnisse, d.h. es werden mehr signifikante Unterschiede aufgedeckt.
Für die Hypothesenfamilie $\{H_o^V, V \subseteq \{1,..,d\}\}$ ist das in Abschnitt 5.3.1 entwickelte sequentielle Verfahren besser geeignet (vgl. das Vorwort in Sonnemann (1981b) bezüglich sequentiell verwerfender Testprozeduren).

5.4 Multiple sequentielle Permutationstests für die Faktorstufen in uni- und multivariaten Versuchsplänen zum multiplen Niveau α

Petrondas & Gabriel (1983) haben demonstriert, daß für Permutationstests bezüglich mehrerer Faktorstufen in einem ANOVA (oder MANOVA-) Testproblem simultane Testprozeduren das multiple Niveau α nicht einhalten. In STPen werden bekanntlich alle Tests von Teilhypothesen mit Hilfe des unter der globalen Nullhypothese gewonnenen Schwellenwertes ausgeführt. Wie in Abschnitt 5.1.4 erläutert, sollten aber die Tests von Teilhypothesen unabhängig von Beobachtungen

unter anderen von der jeweiligen Nullhypothese nicht betroffenen Faktorstufen sein. Damit sind prinzipiell sequentiell verwerfende Teststrategien - neben dem stets moglichen Bonferroni Test - vorzuziehen. Diese sequentiell verwerfenden Tests beginnen mit einer globalen Hypothese und gehen zu immer eingeschrankteren Hypothesen uber bis zum ersten Mal eine Hypothese akzeptiert wird. Es gibt verschiedene solcher Strategien von unterschiedlicher Teststärke, die nun erlautert werden sollen. Im Folgenden seien stets o.B.d.A. durchschnittsabgeschlossene Hypothesenfamilien mit endlich vielen Elementarhypothesen angenommen. Alle wichtigen Eigenschaften dieser Testprozeduren lassen sich am univariaten unverbundenen k-Stichprobenproblem ohne Normalverteilungsannahme erläutern.

(i) Ist man an paarweisen Vergleichen interessiert, so sind neben den $m=k(k-1)/2$ Elementarhypothesen

$H_o^{ij}: \mu_i=\mu_j, i<j=1,..,k$ auch alle Durchschnittshypothesen zu betrachten. Unter diesen gibt es sog. 'Faktorstufen-Teilmengen'-Hypothesen

$$(5.30)\quad H_o^{ijl..}: \mu_i=\mu_j=\mu_l=\dots \quad , 1\le i<j<l<..\le k \text{ und } i<j<l\in\{1,..,k\}$$

und Partitionshypothesen

$$H_o^{ij..,\,lo..}: \mu_i=\mu_j=\dots \text{ und } \mu_l=\mu_o=\dots \quad \text{mit } i<j<..\le k;\ l<o<..\le k$$

Diese Testfamilie $\{H_o^{J(i)}\}$ mit J(i) wie in Abschnitt 5.2 mit 2^k-k+1 Elementen ist nicht komplett (vollstandig), da nicht alle Durchschnitte paarweise verschieden sind. Das verdeutlicht man sich leicht fur k = 3 und

$H_o^{12} \cap H_o^{13} = H_o^{12} \cap H_o^{23}$.

(ii) Beim Vergleich aller k-1 Stufen gegen eine Kontrolle k hat man dagegen eine komplette durchschnittsabgeschlossene Hypothesenfamilie mit den Elementarhypothesen:

$$H_o^i: \ \mu_i = \mu_k \quad i=1,..,k-1$$

(5.31) und den Durchschnittshypothesen

$$H_o^{ijl..}: \ \mu_i = \mu_j = \mu_l = ... = \mu_k \quad i<j<l<..<k \text{ mit } i,j,l \in \{1,..,k-1\} .$$

Für jede der Hypothesen in (5.30) bzw. (5.31) kann man leicht einen exakten Permutationstest zum Niveau α angeben, der nur Permutationen von Beobachtungen unter solchen Faktorstufen als zulässig akzeptiert, auf die sich die jeweiligen Hypothesen der Hypothesenfamilie beziehen.

Für Teilmengen-Hypothesen wählt man wie in (5.1) die Range (Bereichs-) Statistiken oder entsprechende Zähler von ANOVA-F-Statistiken:

Fall (i):
$$Q^{J(i)}(Y) = \max_{j,j' \in J(i)} \{|\bar{Y}_{j.} - \bar{Y}_{j'.}|\} \ , \ i=1,..,n;$$

bzw.
$$\bar{Y}_{j.} = 1/n_j \sum_{l=1}^{n_j} Y_{jl} \ , \ n_{.} = \sum_j n_j$$

(5.32) oder
$$F^{J(i)}(Y) = \sum_{j \in J(i)} (\bar{Y}_{j.} - \bar{Y}^{J(i)}_{..})^2 \qquad \bar{Y}^{J(i)}_{..} = \sum_{j \in J(i)} \sum_{l=1}^{n_j} Y_{jl} / \sum_{j \in J(i)} n_j$$
$$= \sum_{j \in J(i)} \bar{Y}^2_{j.} \ , \text{ da } \bar{Y}^{J(i)}_{..} \text{ permutationsinvariant unter der jeweiligen } H_o^{J(i)}.$$

Fall (ii):
$$Q_k^{J(i)}(Y) = \max_{j \in J(i)} \{|\bar{Y}_{j.} - \bar{Y}_{k.}|\}, \ J(i) \subseteq \{1,..,k-1\} .$$

Anstelle der Teststatistikwerte geht man wie zuvor zu den zugehörigen p-Werten über:

(5.33)
$$p^{J(i)} = |\{\pi \in S_{J(i)} \mid T^{J(i)}(\pi y) \geq T^{J(i)}(y)\}| / |S_{J(i)}|$$

mit
$$T^{J(i)} \equiv Q^{J(i)} \text{ oder } F^{J(i)} \text{ oder } Q_k^{J(i)} .$$

Für Partitionshypothesen bezüglich z.B. 2 Gruppen von Faktorstufen kann man als Teststatistiken wählen:

$$Q^{ij..,\,lo..}(Y) = \max\{Q^{ij..}(Y),\ Q^{lo..}(Y)\}$$

bzw.

$$F^{ij..,\,lo..}(Y) = \max\{F^{ij..}(Y),\ F^{lo..}(Y)\}$$

(5.34) mit den zugehörigen p-Werten:

$$p^{ij..,\,lo..} = |\{\pi \in S_{ij..} \times S_{lo..} \mid T^{ij..,\,lo..}(\pi Y) \geq T^{ij..,\,lo..}(y)\}| \,/\, |S_{ij..}| \cdot |S_{lo..}|$$

mit $T \equiv Q$ oder F.

Für die weiter unten dargestellten sequentiell verwerfenden multiplen Tests kann man auf die explizite Bestimmung dieser p-Werte für Partitionshypothesen verzichten. Nach dem Vereinigungs-Durchschnittsprinzip wird für die Durchschnittshypothesen mit Niveau-α Vereinigungs-Durchschnittstests (vgl. (5.22)) wie in (5.21) entschieden und anstelle der expliziten Ausführung des Abschlußtests sequentiell vorgegangen von der globalen Nullhypothese über die Identität aller k Parameter bis hin zu den jeweiligen Elementarhypothesen.

5.4.1 Vergleiche aller Faktorstufen

Von den verschiedenen möglichen Verfahren, die alle multiple Tests zum multiplen Niveau α seien, soll dasjenige bevorzugt werden, welches die größte oder eine möglichst große Teststärke bezüglich aller Paarvergleiche besitzt (vgl. Welsch, 1977). Im parametrischen Fall (vgl. Ramsey, 1978) ist die sequentielle Prozedur von Peritz, beschrieben in Einot & Gabriel (1975), Begun & Gabriel (1981) und Hommel (1985) sehr geeignet. Für Rangtests demonstrieren Campbell & Skillings (1985) die Überlegenheit dieser sequentiellen Verfahren. Petrondas & Gabriel (1983) zeigen, daß dieses Verfahren auch fur Permutationstests verwendet werden kann.

Folgende multiple Testprozeduren werden dazu verwendet:

(1) Für Hypothesenfamilien bestehend aus Teilmengen-Hypothesen verwendet man das Verfahren von Ryan (1959, 1960, vgl. Hommel (1985)):

Sei

$$J:=\{1,..,k\}\ ,\ |J|=k \quad \text{und} \quad J(j)\subseteq J(i)\ ,\ 2\leq|J(j)|\leq k$$

$$|J(j)|:=k_j \qquad k_i:=|J(i)| \quad ;\ j,i\in\{1,..,n\} \quad .$$

Sei $p^{J(j)}$ definiert wie in (5.33) unabhängig von Faktorstufen aus $J\cap J^c(j)$ mit $J^c(j):=J\setminus J(j)$ dem Komplement von $J(j)$ bezüglich J.

Die kohärente multiple Testprozedur zum multiplen Niveau α lautet dann (vgl. Einot & Gabriel, 1975, Abschnitt 1.7):

(5.35) Für alle J(j) gilt: $$\phi_{J(j)}=\begin{cases}1 & \leq \text{ für } \underline{\text{alle}}\ J(i)\\ p^{J(i)} & \alpha k_i/k \quad \text{mit } J(j)\subseteq J(i)\\ 0 & > \text{ für ein } J(i)\end{cases}$$

Anstelle von $c_\alpha(J(i)) = \alpha\cdot k_i/k$ benutzen Einot & Gabriel die etwas weniger konservativen Schranken $1-(1-\alpha)^{k_i/k}$ $(\geq\alpha k_i/k)$. Das oft verwendete Verfahren von Newman & Keuls mit $c_\alpha(J(i))=\alpha$ hält das multiple Niveau α nicht ein (vgl. Einot & Gabriel, 1975, Abschnitt 1.7; Horn, 1981).

Wie zuvor erwähnt, kann man die Ryan-Prozedur schrittweise ausführen, ohne alle Teilmengen überprüfen zu müssen; sie ist in Abbildung 5.2 dargestellt.

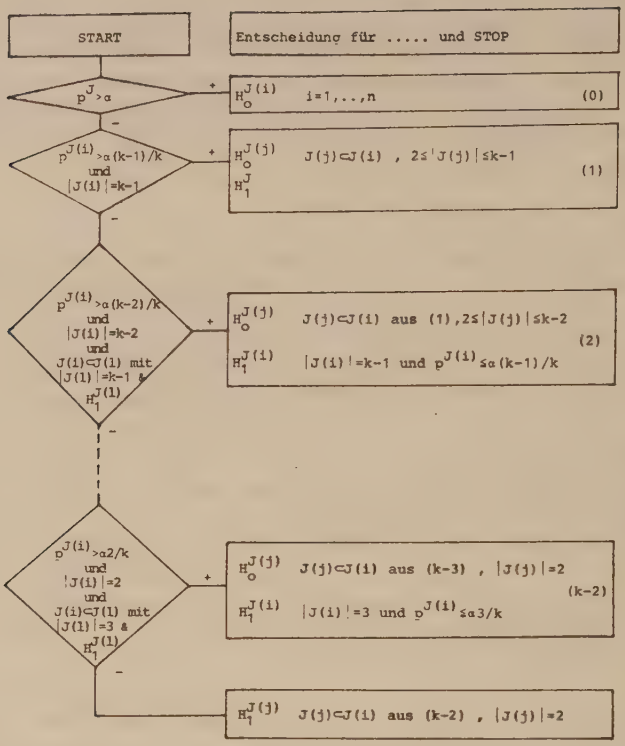


Abb. 5.2.: Sequentieller multipler Permutationstest zum multiplen Niveau α zum Vergleich von $J=\{1,..,k\}$ Faktorstufen nach Ryan (1960) bzw. Einot & Gabriel (1975), wenn man $a \cdot l/k$, $l=k-1,..,2$, ersetzt durch $1-(1-\alpha)^{l/k}$; Definition der $p^{J(i)}$ nach (5.33).

Man muß also nur in den Teilmengen weitertesten, für die sich ein signifikantes Ergebnis gezeigt hat. Für alle anderen J(i) werden auch alle ihre Teilmengenhypothesen akzeptiert. Diese 'Annahme per Implikation' erzeugt Kohärenz der Testentscheidungen. Aufgrund der Struktur von Schritt (1) in der Abbildung, d.h. es wird nur H_1^J selbst akzeptiert, falls mindestens ein $p^{J(i)} > \alpha(k-1)/k$ und $p^J \le \alpha$, kann in diesem Schritt auch α anstelle von $\alpha \cdot (k-1)/k$ gewählt werden (vgl. Welsch, 1977; Ramsey, 1978, Formel (1.3);

Hommel, 1985, S. 39).

(2) Die Hypothesenfamilie bestehe aus allen möglichen Partitionshypothesen bezüglich der k Faktorstufen:

$$(5.36)\qquad H_o^{\nu *} = \bigcap_{i=1}^{g} H_o^{J(i)} \qquad \text{mit } J(i)\cap J(j)=\emptyset \text{ für } i\neq j\in\{1,..,g\}\ ,$$
$$1\leq |J(i)| \leq k\ ,\ \sum_{i=1}^{g} J(i)=k\ ,\ g=1,..,k\ .$$

In dieser Hypothesenfamilie sind nur Hypothesen enthalten, die Gleichheit zwischen Faktorstufen jeweils innerhalb der Partitionen annehmen. Einelementige Hypothesen sind ausdrücklich zugelassen. Jede der Teilmengenhypothesen $H_o^{J(i)}$ kann als Partitionshypothese dargestellt werden, in der alle Partitionen bis auf J(i) einelementig sind. Diese Hypothesen sind trivialerweise wahr. Damit kann man die Hypothesenfamilien bestehend aus Teilmengenhypothesen dadurch vervollständigen (abschließen), daß man für $k\geq 4$ alle nichttrivialen Partitionshypothesen zu den Teilmengenhypothesen hinzunimmt, die außerdem nicht identisch mit irgendeiner der Teilmengenhypothesen sind. Für $k\geq 4$ gibt es $k!/\prod_{i=1}^{g} k_i!$, $k_i\geq 2$ mit $\sum_{i=1}^{g} k_i=k$, verschiedene echte Partitionshypothesen bis auf Vertauschungen ganzer Partitionen. Falls $k_i=k_o$ und $k=g\cdot k_o$ ist, reduziert sich die Anzahl um den Faktor g!. Partitionshypothesen $H_o^{\nu *}$ können direkt wie in (5.34) mithilfe von p_ν oder wie in Petrondas & Gabriel (1983) vorgeschlagen getestet werden anhand folgender p-Werte:

$$(5.37)\qquad p_\nu^* = \begin{cases} p^{J(i)} & \text{falls } H_o^{\nu *}=H_o^{J(i)} \text{ für ein } H_o^{J(i)} \\ 1-\left(\max_{i\in J(i)} \{\sqrt[k_i]{1-p^{J(i)}}\}\right)^k & \text{sonst} \end{cases},$$

d.h. $H_o^{\nu *}$ wird nicht verworfen, falls alle $p^{J(i)}>1-(1-\alpha)^{k_i/k}$ sind.

Kohärenz der Testprozedur wird durch folgendes Vorgehen erzwungen: Eine Partitionshypothese $H_o^{\nu *}$ wird akzeptiert, falls für alle Partitionshypothesen H_o^{l*}, die in $H_o^{\nu *}$ enthalten sind, deren p-Wert nach (5.34) größer als α ist. H_o^{l*} ist in $H_o^{\nu *}$ enthalten, falls jede Partition, auf die sich H_o^{l*} bezieht, in einer Partition unter $H_o^{\nu *}$ enthalten ist.

Dieses Verfahren ist sehr aufwendig, da jeweils alle enthaltenen Partitionen geprüft werden müssen. Das von Peritz vorgeschlagene Verfahren (vgl. Petrondas & Gabriel, 1983, S. 900) ist einfacher auszuführen:

> Eine Partitionshypothese $H_o^{\nu *}$ wird akzeptiert, falls p_l^* nach (5.37) $> \alpha$ für alle Partitionshypothesen mit $H_o^{l*} \subset H_o^{\nu *}$. Für Teilmengenhypothesen $H_o^{J(i)}$ bedeutet das äquivalent: $H_o^{J(i)}$ wird genau dann akzeptiert, wenn für jedes J(j) mit $J(i) \subseteq J(j)$ gilt:
>
> (i) $p^{J(j)} > \alpha \cdot |J(j)|/k$
>
> (5.39) und
>
> (ii) $p^{J(g)} > \alpha \cdot |J(g)|/k$ für jede Teilmenge J(g) im Komplement von J(j), d.h. $J(g) \subseteq J \cap J^c(j)$.

Hier könnten auch wieder in (i) $1-(1-\alpha)^{|J(j)|/k}$ und in (ii) $1-(1-\alpha)^{|J(g)|/k}$ gewählt werden.

Begun & Gabriel (1981) und für Permutationstests Petrondas & Gabriel (1983) entwickeln daraus einen sequentiell verwerfenden multiplen Test, der explizit keine Partitionshypothesen prüft. Für die anhand der Teststatistiken aus (5.32) gewonnenen p-Werte entscheidet man, beginnend mit der globalen Hypothese H_o^J: $\mu_1 = \ldots = \mu_k$ und fortfahrend mit den $H_o^{J(i)}$ mit $|J(i)| = k-1$, usw.:

1. $p^{J(i)} > \alpha \Rightarrow H_o^{J(i)}$ und alle $H_o^{J(j)}$ mit $J(j) \subset J(i)$.

   Es werden also alle Nullhypothesen beibehalten, die auch mit dem Verfahren von Newman & Keuls (vgl. Miller, 1981) akzeptiert werden.

2. $p^{J(i)} \leq \alpha \cdot |J(i)|/k \Rightarrow H_1^{J(i)}$

   Es werden also alle Nullhypothesen verworfen, die auch nach dem Verfahren von Ryan (1960) verworfen werden.

(5.40) 3. $\alpha \cdot |J(i)|/k < p^{J(i)} \leq \alpha$ (sog. kritische Hypothesen).

   Beginnend mit den p-Werten für die größten Teilmengen von Faktorstufen entscheidet man:

   (a) Falls ein $J(j)$ mit $J(j) \subset J^c(i) \cap J$ existiert mit $p^{J(j)} > \alpha \cdot |J(j)|/k \Rightarrow H_o^{J(i)}$ wird akzeptiert oder sie wird auch akzeptiert, falls $J(i) \subset J(j)$ mit $\alpha \cdot |J(j)|/k < p^{J(j)} \leq \alpha$ und $H_o^{J(j)}$, d.h. falls die kritische Hypothese $H_o^{J(j)}$ zu einem früheren Schritt akzeptiert wurde.

   (b) Falls für alle $J(j)$ mit $J(j) \subset J^c(i) \cap J$ und $|J(j)| \geq 2$: $p^{J(j)} \leq \alpha \cdot |J(j)|/k \Rightarrow H_1^{J(i)}$ wird akzeptiert und falls $J(i) \not\subset J(j)$ für ein in einem früheren Schritt kritisches $H_o^{J(j)}$, d.h. falls die kritische Hypothese $H_o^{J(j)}$ in einem früheren Schritt akzeptiert wurde.

Weniger konservativ könnte auch wieder in 3. mit $1-(1-\alpha)^{|J(i)|/k}$ gearbeitet werden. Falls $|J(i)|=k-1$ ist und $\alpha \cdot (k-1)/k < p^{J(i)} \leq \alpha$ (d.h. Fall 3.), wird $H_o^{J(i)}$ akzeptiert, denn das Komplement von $J(i)$ enthält nur ein Element und die zugehörige Hypothese ist deshalb trivialerweise erfüllt. Deshalb kann auf dieser Stufe $\alpha \cdot (k-1)/k$ durch α ersetzt werden.

Testentscheidungen über Partitionen sind implizit im Algorithmus enthalten, denn $H_o^{v^*}$ wird akzeptiert, falls $H_o^{J(j)}$ akzeptiert wird für alle j=1,..,g.

Das Verfahren von Peritz erlaubt die Ablehnung einer Teilmengenhypothese nur, wenn alle Partitionshypothesen, die diese Teilmengenhypothese als Teil enthalten, verworfen werden.

Mit obigem Algorithmus hat man eine - verglichen mit dem Ryan-Verfahren - teststärkere multiple Testprozedur, die das multiple Niveau α im Gegensatz zum Newman-Keuls Verfahren einhält (vgl. Einot & Gabriel, 1975, Abschnitt 1.7 zum Beweis, daß das Newman-Keuls Verfahren das nicht tut).

Im parametrischen Fall mit Normalverteilungsannahme gibt es Untersuchungen dazu, inwieweit man durch dieses kompliziertere Verfahren an Teststärke gewinnt. Einot & Gabriel (1975) berichten von einer Monte-Carlo Studie mit k=5, daß nur für den Fall der F-Statistik (vgl. (5.32)) die Unterschiede in der Teststärke beträchtlich sein können (β(Peritz) > β(Ryan) > β(STP)), während für die Range-Statistik kaum Differenzen vorkommen. Die Autoren vermuten auch, daß für Abweichungen von der Normalverteilung sich keine großen Änderungen in den Relationen ergäben. Begun & Gabriel (1981) geben dagegen Konstellationen von Parameterwerten an, für die das Peritz-Verfahren deutlich größere Teststärke besitzt als Ryan's multiple Testprozedur. Das ist der Fall für mehrere in sich homogene, aber untereinander deutlich verschiedene Teilmengen von Parametern. Dasselbe gilt für ein einzelnes homogenes Cluster von z.B. k-1 Parameterwerten innerhalb aller k Stufen. Zum selben Ergebnis gelangt Ramsey (1978) für den Fall von Parameterwerten, die alle jeweils um denselben Betrag voneinander verschieden sind. Auch für die

Aufdeckung aller tatsächlich verschiedenen Paare von Faktorstufen hat die Peritz-Prozedur mit dem kritischen Niveau α für $|J(i)|=k-1$ die größte Teststärke.

Dieser Effekt der Abschlußprozeduren ist leicht einzusehen. In den oben genannten Fällen ist die Wahrscheinlichkeit groß, daß auch für das Komplement zu einer Teilmenge die Nullhypothese verworfen werden muß.

Der Vor- oder Nachteil dieser Abhängigkeit wird aber von den verschiedenen Autoren unterschiedlich gewichtet. Einot & Gabriel bezeichnen sie als wenig wünschenswert, während Begun & Gabriel den praktischen Vorteil der größeren Teststärke hervorheben. Die Diskussion wird in Abschnitt 6.1.4 wieder aufgenommen. Ansätze zu einem analytischen Vergleich der Prozeduren findet man bei Hommel (1985).

Welsch (1977) hat einen sequentiell verwerfenden multiplen Test zum multiplen Niveau α vorgeschlagen, in dem schrittweise von 'unten' (step-up Procedure), d.h. von den Elementarhypothesen über Paare von Parametern ausgegangen wird. Übertragen auf die p-Werte von Permutationstests lauten die Regeln zur Testentscheidung:

(5.41)
Beginnend mit den $H_o^{J(i)}$ mit $|J(i)|=2$ wird entschieden:
(i) Falls $p^{J(i)} \leq \alpha \cdot |J(i)|/k \Rightarrow H_1^{J(j)}$ für alle $J(j) \supseteq J(i)$.
(ii) Falls $p^{J(i)} > \alpha \cdot |J(i)|/k \Rightarrow H_o^{J(j)}$ für alle $J(j) \subseteq J(i)$.

Anstelle von k-1 kann im Algorithmus k verwendet werden.

Weniger konservativ sind die Entscheidungen, wenn man wie zuvor $1-(1-\alpha)^{|J(i)|/k}$ als kritische Schranken wählt. Welsch vermutet, daß sein Verfahren (5.41) ähnlich gut wie das Verfahren von Peritz ist.

Alle zuvor dargestellten sequentiellen Verfahren sind direkt übertragbar auf andere ein- oder mehrfaktorielle univariate Versuchspläne. Durch die Bedingung, daß stets nur zulässige Permutationen bezuglich der von der jeweiligen Teilhypothese betroffenen Faktorstufen betrachtet werden, sind die exakten Verteilungen und damit die p-Werte stets bei Gültigkeit der jeweiligen Teilhypothese vollständig bestimmt. Dasselbe gilt fur die Tests von analogen multivariaten Hypothesenfamilien mit Elementen

$$(5.42)\qquad H_o^{J(i)} : \mu_l = \bar{\mu}^{J(i)} \quad \text{für alle } l\in J(i) \quad \text{und } \mu_l=(\mu_l^1,..,\mu_l^d)' \text{ , } d>1 \text{ .}$$

Bei der Auswahl einer geeigneten Teststatistik gibt es wieder mehrere Möglichkeiten. Jede der Teststatistiken aus (4.18) ist benutzbar. Alternativ wäre es z.B. auch möglich, nur den ersten Schritt der multiplen Testprozedur für die d abhängigen Variablen aus Abbildung 5.1 zu nutzen und $H_o^{J(i)}$ aus (5.42) zu verwerfen, zu akzeptieren oder als kritisch anzusehen, je nachdem ob

$$(5.43)\qquad p_{J(i)}^{[1]} = 1/|\Omega_N| \cdot |\{\pi\in\Omega_N \mid \min_{l=1,..,d}\{p_l(S_l^{J(i)}(x_N^l))\}\leq p_{(1)}^{J(i)}\}|$$

$$\begin{cases} \leq \alpha\cdot|J(i)|/k \\ > \alpha \\ \in(\alpha\cdot|J(i)|/k,\alpha] \end{cases} \quad \text{ist.}$$

Statt $|J(i)|/k$ könnte auch $1-(1-\alpha)^{|J(i)|/k}$ benutzt werden.

5.4.2 Vergleiche gegen eine Kontrolle

Multiple Vergleiche gegen eine Kontrolle sind u.a. immer dann von Interesse, wenn mehrere experimentelle Bedingungen (z.B. neue Methoden) gegen eine Kontrollbedingung (Standardmethode) verglichen werden sollen, Unterschiede zwischen den experimentellen Bedingungen aber eine untergeordnete Rolle spielen. Häufig sind nur einseitige Tests von Interesse, in denen für alle experimentellen Bedingungen größere/kleinere Werte als die in der Kontrollbedingung angenommen werden. In der Teststatistik (5.32), Fall (ii) könnte dann die Betragsbildung entfallen.

Wie bereits in Abschnitt 5.4 erwähnt, ist die zugehörige Hypothesenfamilie mit den Elementarhypothesen

$$H_o^i: \mu_i = \mu_k \quad (\text{oder } \mu_i \leq \mu_k) \quad , \; i=1,..,k-1$$

durchschnittsabgeschlossen und komplett.

Wie schon in der Hypothesenformulierung angedeutet, sind die interessierenden Elementarhypothesen in der Regel einseitig zu formulieren. Die im Folgenden dargestellten schrittweisen Testprozeduren erlauben aber, wenn die jeweilige Nullhypothese verworfen wird, keine Entscheidung darüber, ob nun $\mu_j > \mu_k$ oder $\mu_j < \mu_k$ für (alle) $j \in J(i)$. Es sind also sog. 'directional errors' (Shaffer, 1980) oder Fehler III. Art möglich. Nur unter zusätzlichen, allerdings häufig erfüllbaren Voraussetzungen an die Verteilungsklassen, aus denen die Beobachtungen stammen (vgl. Shaffer, 1980, Theorem 2) kann man auch die 'Richtungs-Entscheidung' treffen. Wenn alle Hypothesen der Hypothesenfamilie einseitig sind, müssen die von Shaffer formulierten Zusatzbedingungen nicht erfüllt sein. Wie bei Marcus, Peritz & Gabriel (1976, S. 656) sei deshalb (vorsichtshalber) angenommen, daß stets $\mu_i - \mu_k > 0$ gilt

für alle i=1,..,k-1.

Entsprechend kann die Teststatistik (5.32, (ii)) ersetzt werden durch

$$(5.44)\quad Q_k^{J(i)}(Y) = \max_{j\in J(i)} \{\bar{Y}_{j.}\} - \bar{Y}_{k.} \quad , J(i)\subseteq\{1,..,k-1\} .$$

Dabei seien die $2^{k-1}-1$ Hypothesen der kompletten, durchschnittsabgeschlossenen Hypothesenfamilie wieder mit J(i), $i=1,..,2^{k-1}-1$ gekennzeichnet. Wie für die Hypothesenfamilie aus Abschnitt 5.4.1 sind bei Ausführung der exakten Niveau-α Tests für die einzelnen Teilhypothesen stets nur Permutationen der Beobachtungen unter den in der Teilhypothese erfaßten Faktorstufen zulässig. Damit ist das sequentielle Verfahren von Marcus, Peritz & Gabriel (1976) für den parametrischen Fall nicht direkt übertragbar, da im Fall von Permutationstests keine monotone Beziehung zwischen den α-Quantilen von studentisierten Spannweiten für unterschiedliche Faktorstufenanzahlen besteht. So kann das obere $(1-\alpha)$-Quantil von $Q_k^{J(i)}$ mit $|J(i)|=2$ größer als das Quantil der Teststatistik für die Globalhypothese ausfallen. Darum müssen wie im Fall aller Vergleiche auch hier lediglich schrittweise Verfahren, die nur die p-Werte für die einzelnen Hypothesen der Hypothesenfamilie benötigen, Verwendung finden.

Wenn man sich nur für die Elementarhypothesen interessiert, ist stets die Anwendung des Bonferroni-Holm Verfahrens auf die zu den k-1 unverbundenen (einseitigen) Zweistichprobenproblemen gehörenden p-Werte mit den nominellen α-Niveaus α/l, $l=k-1,..,1$ möglich.

Von größerer Teststärke ist das in Abb. 5.2 bereits dargestellte Verfahren der α-Niveau Vergabe nach Ryan bzw. Einot & Gabriel. Da im Fall der Vergleiche gegen eine Kontrolle nur k-1 Elementarhypothesen vorhanden sind, ist in dem Flußdiagramm durchgehend k durch k-1 zu ersetzen. Da verglichen mit dem Verfahren für alle Teilmengen von Faktorstufen die $2^{k-1}-1$ Teilmengenhypothesen, die die Kontrollbedingung enthalten, berücksichtigt werden müssen, ist dieses Verfahren sicher noch nicht optimal. Die Struktur der Hypothesenfamilie impliziert z.B., daß es keine Partitionshypothesen, sondern nur Teilmengenhypothesen unter Einschluß der Kontrollbedingung gibt. In diesem Fall ist es, im Unterschied zu der Hypothesenfamilie aus dem vorhergehenden Abschnitt erlaubt, sequentiell beginnend mit der Globalhypothese bis zu den Elementarhypothesen aus (5.31) vorzugehen und stets die p-Werte zu den Statistiken aus (5.44) mit demselben multiplen Niveau α zu vergleichen. Die sequentielle multiple Testprozedur läßt sich somit folgendermaßen formulieren:

(5.45) Beginnend mit der globalen Hypothese $H_o^{J(1)}: \mu_1=\ldots=\mu_k$ und fortlaufend mit den $H_o^{J(i)}, |J(i)|=k-1$, und $k\in J(i)$ usw. bis hin zu den $H_o^{J(i)}$ mit $|J(i)|=2$ und $k\in J(i)$ entscheidet man:

1. $p^{J(i)}>\alpha \Rightarrow H_o^{J(i)}$ und alle $H_o^{J(j)}$ mit $J(j)\subset J(i)$.
2. $p^{J(i)}\leq\alpha \Rightarrow H_1^{J(i)}$.

Wichtig für die Einhaltung des multiplen α-Niveaus ist die schrittweise, Kohärenz erzwingende Vorgehensweise.

Zum Beweis überlegt man ähnlich wie bei Einot & Gabriel (1975) Folgendes: Die interessierende Größe ist die maximal mögliche Wahrscheinlichkeit für einen Fehler 1. Art. Dazu ist es ausreichend, die fälschliche Ablehnung der Null-

hypothese bezuglich der größten homogenen Teilmenge $J(i^*)$ von Faktorstufen unter Einschluß der Kontrollbedingung k zu betrachten. Wegen der die Kohärenz der Testentscheidungen erzwingenden Vorgehensweise wird, falls eine kleinere homogene Teilmenge fälschlicherweise als nicht homogen angesehen wird, auch die Nullhypothese bezüglich dieser großeren Teilmenge von Faktorstufen fälschlicherweise zum Niveau α verworfen. Diese Wahrscheinlichkeit einer fälschlichen Ablehnung wird am größten, wenn alle anderen wahren Faktorstufen-Parameter außerhalb dieser Teilmenge 'sehr weit' von der Kontrollbedingung getrennt sind. Die Wahrscheinlichkeit für einen Fehler 1. Art bezüglich $J(i^*)$ kann aber das nominelle Fehlerniveau α trivialerweise nicht überschreiten, da die Permutationstests für die Hypothesen der Hypothesenfamilie alle Niveau-α Tests sind. Wie in Petrondas & Gabriel (1983, Kap. 3.2) begründet, ist für eine durchschnittsabgeschlossene Hypothesenfamilie die Wahrscheinlichkeit für einen Fehler 1. Art $\leq \max\{\alpha_{J(i)} | i=1,..,2^{k-1}-1\}$. Alle nominellen Niveaus für die Teilhypothesen sind aber gerade $\alpha_{J(i)}=\alpha$.

5.5 Simultane Permutationstests für die Faktorstufen in uni- und multivariaten Versuchsplänen zum globalen Niveau α

Wie zuvor schon erläutert, ist es mit Permutationstests nicht moglich, simultane Testprozeduren für Hypothesenfamilien bezüglich Faktorstufen zu konstruieren, in denen die Teststatistikwerte für alle Teilhypothesen gegen einen exakten kritischen Wert verglichen werden. Für Teilhypothesen bezüglich Faktorstufen ist stets nur eine Untergruppe

der unter der globalen Nullhypothese bezuglich aller Faktorstufen zulässigen Permutationen zu betrachten. Diese Untergruppe beinhaltet jeweils lediglich nur zulässige Permutationen unter Gultigkeit der Teilhypothese.

Das in Abschnitt 5.1.2 kurz erläuterte Vorgehen von Foutz (1981, 1984) umgeht obiges Problem durch die Kontrolle ausschließlich des globalen α-Niveaus und aller marginalen α - Niveaus für alle Teilhypothesen der Hypothesenfamilie. Das marginale α-Niveau für eine Teilhypothese ist die Wahrscheinlichkeit, bei Gültigkeit der globalen Nullhypothese die Teilhypothese zu verwerfen. Folglich werden die Werte der Teststatistiken bezüglich der Teilhypothesen für alle unter der globalen Nullhypothese zulässigen Permutationen bestimmt. Für diese gemeinsame r-dimensionale exakte Verteilungsfunktion der r Teststatistiken unter $H_o = \bigcap_{i=1}^{r} H_o^i$ kann man mit dem in Abschnitt 5.1.2 beschriebenen Algorithmus mit den Schritten (i)-(v) dort für das Beispiel aller r=k(k-1)/2 Paarvergleiche im unverbundenen k Stichprobenproblem die exakten marginalen Schwellenwerte bestimmen und mit den beobachteten Prüfgrößen vergleichen. Wie in (5.5) für diesen Spezialfall angegeben, kann man dann auch exakte marginale p-Werte leicht bestimmen.

Das Konstruktionsverfahren für dieses simultane Testverfahren ist sehr allgemein und kann direkt auf Hypothesenfamilien bezüglich aller vorgestellten uni- und multivariaten Permutationstests verallgemeinert werden. Foutz et al. (1985) selbst stellen Tests für die Analyse von Wachstumskurven dar, die unter verschiedenen ein- oder mehrfaktoriellen experimentellen Bedingungen (CR-k, RB-k) beobachtet wurden. In mehrfaktoriellen Plänen sind jedoch u.U. zuerst 'störende' Effekte zu eliminieren. In verte-

lungsfreien MANOVA-Testproblemen (z.B. pCR-k) konnen auch Hypothesenfamilien der Form

$$\{H_o^{ii',d} : \mu_{i,d} = \mu_{i',d} \;,\; i \neq i' \in \{1,..,k\},\; d \in \{1,..,p\}\}$$

mit dem Verfahren überprüft werden.

In Abschnitt 8.4.5 ist ein Computerprogramm beschrieben, das auf obigem Algorithmus basiert.

Diskussionswürdig ist bei dem Ansatz von Foutz allerdings, daß das simultane Testverfahren zwar das globale i.a. aber nicht das multiple Niveau α einhalt. Wie in den Beispielen von Petrondas & Gabriel (1983) in anderem Zusammenhang demonstriert. kann der beobachtete p-Wert für jede der Teilhypothesen bei der auf diese Teilhypothese eingeschränkte zulassige Permutationen den p-Wert bezüglich aller unter der globalen Nullhypothese zulässigen Permutationen je nach Konstellation der Werte unter den anderen Faktorstufen über- oder unterschreiten. Die Bedingung in Sonnemann (1982, Bed. 3.4) muß also nicht erfüllt sein. Es ist nur die schwächere Bedingung erfullt, daß alle marginalen α-Niveaus bei Gultigkeit der globalen Nullhypothese eingehalten werden. Folgt man der Argumentation von Welsch (1977), ist der Bezug auf die globale Nullhypothese für multiple Tests ungeeignet. Wenn multiple Tests zur Exploration benutzt werden, ist die globale Nullhypothese über die Gleichheit aller Parameter kein besonders geeigneter Bezugsrahmen.

5.6 Multiple Permutationstests für Hypothesenfamilien bezüglich Komponenten multivariater Beobachtungen und bezüglich Faktorstufen

Bei Ablehnung der Globalhypothese in einem multivariaten Testproblem ist es oft sinnvoll, die genaueren Gründe für dieses Verwerfen anhand multipler (Permutations-) Testverfahren zu bestimmen. Je nach Fragestellung sind unterschiedliche Hypothesenfamilien von Interesse. Für den Normalverteilungsfall hat Gabriel (1969b) geeignete, in verschiedener Hinsicht optimale simultane Testverfahren dargestellt. Die Verbesserungen bestimmter simultaner univariater Testprozeduren durch sequentiell verwerfende Verfahren (Einot & Gabriel, 1975; Marcus, Peritz & Gabriel, 1976; Begun & Gabriel, 1981; Sonnemann, 1982) erfolgten jedoch erst später. Letztere Verfahren sind direkt auf den parametrischen multivariaten Fall übertragbar (vgl. z.B. Schiller, 1981). Für Permutationstests ist das Vorgehen je nach Hypothesenfamilie unterschiedlich. Zur Darstellung aller wichtigen Aspekte ist es wieder ausreichend, das d-variate unverbundene k-Stichprobenproblem zu betrachten. Auch wenn es sich um ein d-variates Testproblem handelt, sei im Folgenden angenommen, daß Hypothesen bezüglich der einzelnen Komponenten oder bezüglich Teilmengen der d abhängigen Variablen von Interesse sind.

(i) $\{H^{G,V}: \mu_i^V=\mu^{G,V},\ i\in G\subset\{1,..,k\}, |G|>1;\ V\subset\{1,..,d\}, |V|\geq 1\}$

Die Hypothesenfamilie entspricht der Familie Ω in Gabriel (1969b, S. 70). Die $dk(k-1)/2$ Elementarhypothesen sind

$$H_o^{ij,l}: \mu_i^l=\mu_j^l\ ,\ i\neq j\in\{1,..,k\},\ l\in\{1,..,d\}\ .$$

Wenn man nur an den Elementarhypothesen interessiert ist, kann man $d \cdot k(k-1)/2$ Permutationstests für das unverbundene Zweistichprobenproblem ausführen und die p-Werte nach der Bonferroni- oder der Bonferroni-Holm Prozedur beurteilen. Sind zusätzlich Hypothesen $H_o^{G,V}$ mit $|G|>2$, $|V|>1$ von Interesse und will man ein teststärkeres Verfahren als die gerade genannten verwenden, sind folgende Strategien geeignet. Betrachtet man das detailliert dargestellte Beispiel in Petrondas & Gabriel (1983, Tab. 3), sieht man u.a., daß die Abschluß-Prozedur von Peritz (vgl. Abschnitt 5.4.1) dann für die Vergleiche von Paaren von Faktorstufen teststärker ist als es die Bonferroni-Holm Prozedur wäre, denn der kleinste p-Wert für Paarvergleiche ist

$.029 > .10/6\ (6-1)/2 = .0067.$

Wie in (5.40) bestimmt man $p_{J(i)}^{[1]}$ aus (5.43) für die d-dimensionalen Teilmengenhypothesen bis 'hinunter' zu den d-dimensionalen Paarvergleichen.

(5.46) Falls nach der Peritz-Prozedur Paarvergleichshypothesen abgelehnt werden, wird für diese der sequentielle multiple Test aus Abb. 5.1 ausgeführt, um die abhängigen Variablen zu identifizieren, für die die multiplen p-Werte $p_{J(i)}^{[1]}$ mit $|J(i)|=2$ kleiner oder gleich $\alpha \cdot 2/k$ sind.

Auf dem Wege dahin kann für die abgelehnten Hypothesen uber d-dimensionale Teilmengenhypothesen mit mehr als zwei Elementen stets für multiple Tests bezuglich der abhängigen Variablen das Verfahren aus Abb. 5.1 mit kritischen p-Werten $\alpha \cdot |J(i)|/k$ Verwendung finden.

Obiges Verfahren hat den Vorteil, nur univariate Permutationstests zu benotigen. Für die Entscheidungen bezüglich Hypothesen über eindimensionale Parameter werden ja multiple

p-Werte anhand der univariaten Permutationsteststatistiken aus Abschnitt 3.1.5 (iii) ermittelt.

Bisher ist es ungeklärt, ob diese Entscheidungen über multivariate Hypothesen anhand univariater Teststatistiken eine größere Teststärke besitzen als die in Abschnitt 4.5.3 vorgestellten multivariaten Statistiken gebildet aus Funktionen der Eigenwerte von Hypothesen- und Fehlermatrizen. Die zu ihnen gehörenden multivariaten p-Werte $\tilde{p}_{J(i)}$ könnte man alternativ zu den $p^{[1]}_{J(i)}$ aus (5.43) verwenden und ansonsten wie dort beschrieben vorgehen.

(ii) $\{H_o^{G,a}: a'\mu_i = a'\mu^G, |G|>1, a \in R^d \text{ und } a \neq 0\}$

und

$\{H_o^{G,A}: A'\mu_i = A'\mu^G, |G|>1, A \in \{C_{d\times d-1} | 1_d C = 0, 2 \leq rg(C) \leq d-1\}$

(Hypothesenfamilie $\tilde{\tilde{\Omega}}$ nach Gabriel, 1969b)

Mit dem Peritz-Verfahren werden die Teilmengenhypothesen bezüglich der d-dimensionalen abhängigen Variablen mit Permutationstests unter Verwendung von MANOVA-Teststatistiken $\psi_{J(i)}(Y)$ getestet. Für die abgelehnten Teilmengenhypothesen werden die exakten Verteilungen erstellt und die $(1-\alpha|J(i)|/k)$-Quantile oder deren zum oberen Ende der Verteilung hin nächsten erreichbaren Quantile bestimmt. Für Teststatistiken mit kritischen Werten am unteren 'Ende der Verteilung' sind es entsprechend die $\alpha|J(i)|/k$ -Quantile. Dann vergleicht man die interessierenden Linearkombinationen oder Kontraste in den abhangigen Variablen $\psi_{J(i)}(a'y)$ bzw. $\psi_{J(i)}(A'y)$ mit diesen Quantilen.

Theoretisch ist auch folgendes aufwendige Vorgehen denkbar: Für alle 2^k-k-1 Teilmengenhypothesen wird die exakte Permutationsverteilung von $\psi_{J(i)}(y)$ erstellt oder zumindest der

obere/untere α-te Anteil. Man vergleicht $\psi_J(a'y)$, $J=\{1..,k\}$ bzw. $\psi_J(A'y)$ mit dem $(1-\alpha)$-Quantil (bzw. dem α-Quantil je nach Teststatistik) der exakten Verteilung unter der globalen Nullhypothese. Liegt der Wert über (unter) dem Quantil, wird mit dem Peritz-Verfahren und $\psi_{J(i)}(a'y)$ bzw. $\psi_{J(i)}(A'y)$ in den exakten Verteilungen für die Teilmengenhypothesen sequentiell mit dem Algorithmus (5.40) weitergetestet. In dem Algorithmus muß jeweils nur der $p^{J(i)}$-Wert durch den Teststatistikwert $\psi_{J(i)}(a'y)$ und das adjustierte α-Niveau durch das zugehörige $(1-\alpha)$-Quantil ersetzt werden.

(5.47)

1. $\psi_{J(i)}(a'y) < \psi^{J(i)}_{1-\alpha}(y) \Rightarrow H_o^{J(i)}$ und <u>alle</u> $H_o^{J(j)}$ mit $J(j)\subset J(i)$
2. $\psi_{J(i)}(a'y) \geq \psi^{J(i)}_{1-\alpha\,|J(i)|/k}(y) \Rightarrow H_1^{J(i)}$
3. $\psi^{J(i)}_{1-\alpha\,|J(i)|/k}(y) > \psi_{J(i)}(a'y) \geq \psi^{J(i)}_{1-\alpha}(y)$:

   (a) Falls ein $J(j)$ mit $J(j)\subset J^c(i)\cap J$ existiert mit $\psi_{J(j)}(a'y) < \psi^{J(j)}_{1-\alpha\,|J(j)|/k}(y) \Rightarrow H_o^{J(i)}$ wird akzeptiert oder auch falls $J(i)\subset J(j)$ mit $\psi^{J(j)}_{1-\alpha\,|J(j)|/k}(y) > \psi_{J(j)}(a'y) \geq \psi^{J(j)}_{1-\alpha}(y)$ <u>und</u> $H_o^{J(j)}$ gilt.

   (b) Falls für alle $J(j)$ mit $J(j)\subset J^c(i)\cap J$ und $|J(j)|\geq 2$ $\psi_{J(j)}(a'y) \geq \psi^{J(j)}_{1-\alpha\,|J(j)|/k}(y) \Rightarrow H_1^{J(i)}$ wird akzeptiert und falls $J(i)\not\subset J(j)$ für eine in einem früheren Schritt 'kritische' Hypothese $H_o^{J(j)}$.

Anstelle von $a'y$ kann entsprechend überall auch $A'y$ gesetzt werden.

(iii) $\{H_o^{\hat{c},V}: \sum_{i=1}^{k} c_i\mu_i^V=0, \sum_{i=1}^{k} c_i=0, \sum_{i=1}^{k}|c_i|\neq 0;\ V\subset\{1,..,d\}, |V|\geq 1\}$

(Hypothesenfamilie $\mathcal{S}^*$ nach Gabriel)

Gemäß der Argumentation von Petrondas & Gabriel (1983) muß man zusätzlich fordern, daß <u>alle</u> c_i von Null verschieden sind, da sonst die Menge zulässiger Permutationen unter der

jeweiligen Teilnullhypothese eingeschränkt ist.

Man bestimmt zuerst die exakte Verteilung von $\psi(y)$ unter der globalen Nullhypothese. Für einen bestimmten linearen Kontrast charakterisiert durch $g^* := (c_1^*, .., c_k^*)' \in \hat{G}$ wird

$$\psi_{g^*}^{V}(y) = \psi((\underline{S}_e^V + \underline{S}_h^{V,g^*})^{-}(y)\underline{S}_h^{V,g^*}(y))$$

bestimmt und mit dem $(1-\alpha)$-Quantil (bzw. α-Quantil) verglichen. Die Teilhypothese in dieser simultanen Testprozedur wird verworfen, falls $\psi_{g^*}^{V}(y) \geq \psi_{1-\alpha}(y)$ (bzw. $\leq \psi_{\alpha}(y)$) ist. Dieses Verfahren wird i.a. nicht sehr teststark sein.

Falls nur eine feste Zahl von Kontrasten geprüft werden soll, ist folgendes Vorgehen vorzuziehen. Sei

$$(5.48) \quad F_{g^*}(Y) = \sum_{i=1}^{k} c_i^* n_i \bar{Y}_{i.}^2$$

die univariate Teststatistik für einen der m linearen Kontraste für jede der d abhängigen Variablen. Hier kann die Annahme, daß alle c_i^* von Null verschieden sein müssen, fallengelassen werden.

Ein linearer Kontrast in den d abhängigen Variablen wird als signifikant angesehen, wenn

$$(5.49) \quad p_{g^*}^{[1]} = 1/|Q_{n.}| \cdot |\{\pi \in Q_{n.} | \min_{1 \leq j \leq d} p_j(F_{g^*}(\pi y^j))\} \leq p_{(1)}\}| \leq \alpha/m \quad .$$

$Q_{n.}$ wurde anstelle von $S_{n.}$ geschrieben um anzudeuten, daß bei Kontrasten mit einigen $c_i^*=0$ nur über die Faktorstufen permutiert wird, für die von Null verschiedene Gewichte vorliegen. Diesen Bonferroni-Test kann man noch verbessern, wenn man die $p_{g^*}^{[1]}$ der Größe nach ordnet und nach dem Bonferroni-Holm Verfahren sukzessive mit α/m, $\alpha/(m-1)$, usw. vergleicht.

Falls $p_{g^*}^{[1]}$ zur Ablehnung der zugehörigen Hypothesen führt, kann der multiple Test aus Abb. 5.1 verwendet werden, um bezüglich aller Teilmengen von abhängigen Variablen zu

testen. Die $p^{[1]}_{*,g}$, $l>1$, werden dann jeweils mit α/m, bzw. mit dem zugehörigen $\alpha/(m-\gamma)$ verglichen.

(iv) $\{H_o^{G,a}: \sum_{i=1}^{k} c_i a'\mu_i = 0, \sum_{i=1}^{k} c_i = 0, \sum_{i=1}^{k} |c_i| \neq 0, a \in R^d, a \neq 0\}$ und analog für A

(Hypothesenfamilie $\tilde{\Omega}^*$ nach Gabriel)

Wiederum müssen zusätzlich alle c_i von Null verschieden sein. Man bestimmt die exakte Verteilung von $\psi(y)$ unter der globalen Nullhypothese und vergleicht die Teststatistikwerte unter allen Teilhypothesen, d.h. Kontrasten bzgl. aller Faktorstufen für Linearkombinationen oder Kontrasten von abhangigen Variablen

$$\psi(((\underline{S}_e + \underline{S}_h^{g*})(a'y))^{-}\underline{S}_h^{g*}(a'y))$$

gegen das $(1-\alpha)$-Quantil (bzw. das α-Quantil) der Verteilung einer der MANOVA-Statistiken unter dem Permutationsmodell. Ob auch im Fall von Permutationstests die Maximalwurzelkriterien die größte Teststärke (vgl. Krishnaiah, Mudholkar & Subbaiah, 1980) besitzen, ist bisher nicht untersucht.

Auch für Vergleiche gegen eine Kontrolle ist eine direkte Übertragung auf das analoge multivariate Problem möglich. Ebenso können die Verfahren für die zuvor beschriebenen Hypothesenfamilien (i), und (ii) ohne Probleme übertragen werden, wenn man dort stets die Peritz-Entscheidungsprozedur durch die einfachere Newman-Keuls-Prozedur mit den Niveaus $\alpha_i = \alpha$ ersetzt. Die Hypothesenfamilien (iii) und (iv), die allgemeine Kontraste enthalten, entsprechen nicht den Vergleichen gegen eine Kontrolle.

6. TESTSTÄRKE UND ROBUSTHEIT VON PERMUTATIONS- UND RANDOMISIERUNGSTESTS

6.1 Probleme bei der Bestimmung der Teststärke

6.1.1 'Optimale' Permutationstests

Von einem 'guten' statistischen Testverfahren wird man nicht nur verlangen, daß es ein (exakter) Test zum Niveau α ist (vgl. Abschnitt 2). Neben dem Einhalten des α-Niveaus soll dieses Testverfahren bestimmte Abweichungen von der Gültigkeit der interessierenden Nullhypothese mit möglichst großer Wahrscheinlichkeit aufzeigen, d.h. die Teststärke soll möglichst groß, u.U. maximal sein bzw. der Fehler 2. Art (β-Fehler), die Nullhypothese fälschlicherweise zu akzeptieren, soll möglichst klein (minimal) sein.
Für viele (parametrische) Fragestellungen lassen sich in dieser entscheidungstheoretischen Formulierung eines statistischen Testproblems (Neyman-Pearson Formulierung, vgl. z.B. Witte, 1980, Kap. 4.1) (gleichmäßig) teststärkste Testverfahren angeben. Robinson (1979) hat gezeigt, daß man ohne Bezug auf den entscheidungstheoretischen Hintergrund des Hypothesentestens die Methoden von Neyman und Pearson benutzen kann, um zu 'optimalen', d.h. (gleichmäßig) sensitivsten Signifikanztests zu gelangen.
Nur für den Fall einer Lagealternative im Normalverteilungsfall oder für eine vollständig spezifizierte Alternative ist es bisher z.B. für das unverbundene bzw. verbundene Zweistichprobenproblem gelungen nachzuweisen, daß der Randomisierungstest der (gleichmäßig) beste unverzerrte Test für die nichtparametrische Hypothese aus 3.1.5 (i) bzw. (ii) ist

(vgl. Lehmann & Stein, 1948, 1949; Lehmann, 1959, Kap. 5.10; Pratt & Gibbons, 1981, S. 222 ff).
Auch die Ergebnisse von Hoeffding (1952), die zeigen, daß in den betrachteten Fällen der zum ANOVA-Test mit Normalverteilungsannahme analoge Permutationstest asymptotisch (d.h. für wachsende Stichprobenumfänge) dieselbe Teststärke gegen Alternativen des Modells mit Normalverteilung besitzt (d.h. eine asymptotische relative Effizienz ARE=1), sind lediglich von theoretischem Interesse.

Schon Scheffe (1959, Fußnote 28, S. 323) weist darauf hin, daß es viel interessanter wäre, die Teststärke des üblichen F-Tests gegen Alternativen zu bestimmen, die im Randomisierungsmodell erlaubt sind. Aber auch ca. 25 Jahre später wird das Problem optimaler Permutations- und Randomisierungstests noch von Gabriel & Hall (1983, S. 828) als 'second order problem' bezeichnet. Das ist vielleicht auch ein Grund dafür, daß in dem Buch von Edgington (1980) der Begriff Teststärke gar nicht vorkommt.

6.1.2 Unmöglichkeit einer apriori Festlegung der Teststärke

In der parametrischen ANOVA- oder MANOVA-Testsituation ist es möglich, vor Ausführung eines Experimentes die Teststärke als Funktion von Fehler 1. Art, angenommener Effektstärke und festgelegtem (Gesamt-) Stichprobenumfang zu ermitteln (vgl. z.B. Bredenkamp, 1980; Hager & Westermann, 1983a; Hays, 1963; Henning & Muthig, 1979). Es gibt auch drei weitere Möglichkeiten, eine der vier obigen Größen als Funktion der restlichen drei darzustellen (vgl. Hager &

Westermann, 1983a, Kap. 10.2). Mit den Tabellen von Cohen (1977) und den Vorschlägen von Bredenkamp (1980) lassen sich die notwendigen Berechnungen sowohl für multifaktorielle univariate Experimente wie auch für multivariate Experimente besonders bei Verwendung des Pillai'schen Spurkriteriums ohne große Schwierigkeiten ausführen. Entscheidend ist aber neben der üblichen Annahme von additiven Effekten im linearen Modell die Gültigkeit der Normalverteilungsvoraussetzung. Ohne Spezifizierung der Verteilung der Fehler- bzw. Beobachtungsvariablen ist eine exakte Teststärkebestimmung bei der Planung eines Experimentes nicht möglich. Die Annahme, daß die Beobachtungen eine Zufallsstichprobe aus einer Grundgesamtheit mit bekannter (Normal-) Verteilung darstellen, ist also notwendig. Insofern ist der Vorschlag von Bredenkamp (1980, S. 62), für die Festlegung des notwendigen (endlichen) Stichprobenumfangs bei Verwendung des nichtparametrischen Kruskal-Wallis Tests den entsprechenden Umfang für den ANOVA-Fall dividiert durch die asymptotische relative Effizienz gegen die Normalverteilungs-Alternative heranzuziehen, lediglich eine heuristische, approximative Lösung.

Bei Permutationstests sind ganze Klassen von Verteilungen, von denen nur gewisse Symmetrie- oder Vertauschbarkeitseigenschaften vorausgesetzt werden, zugelassen (vgl. die Abschnitte 3 und 4 für die einzelnen linearen Modelle). Permutationstests sind gemäß ihrer Konstruktion bedingte (konditionale) Tests und damit abhängig von den tatsachlich im Experiment realisierten Werten der abhängigen Variable(n). Genauso wie es keine Vertafelung der Teststatistikwerte bei Gültigkeit der jeweiligen Nullhypothese geben kann, ist auch die Verteilung der Teststatistiken selbst bei

Gültigkeit und Bekanntheit einer bestimmten, festen (Lage-) Alternative prinzipiell nicht vor Ausführung des Experiments bestimmbar (vgl. die Vermutung von Bredenkamp, 1980, S. 20). In Randomisierungstests, für die außer einer bestimmten, je Versuchsplan spezifischen Form der zufälligen Zuteilung der Beobachtungseinheiten zu experimentellen Bedingungen keine Zufallsvariablen notwendig sind, gilt obige Aussage 'erst recht' (vgl. Lehmann, 1975, S. 190). Die Teststärke kann nur konditional zu den 'unit values' der jeweiligen Beobachtungseinheiten eines bestimmten Experimentes bestimmt werden (vgl. Easterling, 1975a, b; Lehmann, 1975, S. 57; Puri & Sen, 1970, S. 70, 76f und Abschnitt 1). Auch die unit values (Einheitenwerte) sind in einem Experiment nicht bekannt, sondern nur unter der Annahme konstanter Effekte (vgl. Abschnitt 6.1.3) je Versuchsbedingung berechenbar bzw. schätzbar.

6.1.3 Unmöglichkeit der Festlegung einer durchschnittlichen Teststärke

Im parametrischen ANOVA-Fall läßt sich mit Hilfe der nichtzentralen F-Verteilung (bzw. nichtzentralen χ^2-Verteilung, vgl. Bredenkamp, 1980, S. 53f) der (Mindest-) Stichprobenumfang schätzen, der benötigt wird, um einen Effekt eines Faktors in einem ANOVA-Modell mit einer bestimmten Teststärke aufgrund eines für ein bestimmtes Niveau α durchgeführten F-Tests aufzudecken. Für die praktische Ausführung sind z.B. bei Cohen (1977) sowie Hager & Möller (1983) Tabellen, bei Kirk (1968) Nomogramme angegeben. Der

Nichtzentralitätsparameter ist eine Funktion des Maßes für die durch das lineare Modell aufgeklärte Varianz (vgl. z.B. Bredenkamp, 1980, S. 54). In der einfaktoriellen ANOVA ist der Nichtzentralitätsparameter im wesentlichen eine Funktion von $\sum_{i=1}^{k} (\mu_i - \mu)^2$, d.h. von der mittleren quadratischen Abweichung der Faktorstufenmittel vom Gesamtmittel und der Fehlervarianz (vgl. Hager & Westermann, 1983a, Kap. 9.3.1). Spezifiziert man lediglich den relativen Anteil der Effektvarianz an der Gesamtvarianz, ist eine explizite Kenntnis der Fehlervarianz vor der Versuchsdurchführung überflüssig. Allerdings sind dann auch nur relative mittlere Differenzen mit einem Test nachweisbar. Ohne die Parameterwerte selbst angeben zu müssen, ist damit die Teststärke vor Ausführung des Experiments spezifizierbar und der Stichprobenumfang bestimmbar. Es ist also z.B. für den einfaktoriellen Zufallsplan auch nicht nötig anzugeben, wie groß der Effektunterschied Δ zwischen zwei der k Faktorstufen minimal sein soll. Man kann zwar zeigen, daß die Effektstärke am kleinsten ist, wenn zwei der k Faktorstufen sich um Δ unterscheiden und alle anderen Effektparameter gleich dem arithmetischen Mittel aus diesen beiden Stufen sind (vgl. Cohen, 1977, S. 277; Kirk, 1968, S. 109f; Scheffe, 1959, Kap. 3.3). Aufgrund des inversen Zusammenhangs zwischen Effektgröße und benötigtem Stichprobenumfang ergibt sich aber bei anderer Konstellation der Effektparameter höchstens eine Überschätzung des erforderlichen Stichprobenumfangs. Bredenkamp (1980, S. 85ff) zeigt, daß in der MANOVA bei Verwendung des Pillai'schen Spurkriteriums dieselben Tabellen von Cohen wie für die ANOVA verwendet werden können.

Für Permutations- und Randomisierungstests ist die exakte Verteilung der Teststatistik unter der jeweils interessierenden festen (Lage-) Alternative vor Ausführung des Experimentes trivialerweise nicht bekannt. Da nicht Normalverteiltheit der Beobachtungen gefordert wird sondern eine große Klasse von Verteilungen zugelassen ist, sind prinzipiell keine Tabellen und Nomogramme, insbesondere keine solchen, die im wesentlichen nur von der (durchschnittlichen) Effektstärke abhängen, angebbar.
Man ist gezwungen, feste additive Effekte, die jeweils für alle Beobachtungen unter einer Faktorstufe(nkombination) identisch sein müssen, für jede Faktorstufe(nkombination) zu spezifizieren. Im einfaktoriellen Blockplan muß man z.B. annehmen, daß für jeden der n Blöcke die Abweichungen der individuellen Parameter θ_{ij} (vgl. Abschnitt 3.1.4) von ihrem 'Blockmittel' $\bar{\theta}_{i.}$ konstant sind, und man muß für jede Faktorstufe $j=1,..,k$ den Wert dieser Abweichung $\theta_{ij} - \bar{\theta}_{i.} =: \delta_j$ spezifizieren (vgl. z.B. Gabriel & Hall, 1983, S. 828). Das Vorgehen wird in den Abschnitten 6.2 und 6.3 näher erläutert.

6.1.4 Teststärke von multiplen Testprozeduren

Genauso wie es unterschiedliche Möglichkeiten gibt, verschiedene Typen von Fehlern 1. Art für multiple Testprozeduren zu definieren (vgl. z.B. Kirk, 1968, Kap. 3.3; Miller, 1981, Kap. 1), sind auch verschiedene Ansätze zur Definition von Fehlern 2. Art und der Festlegung von Effektstärken in Abhängigkeit von der Familie zu testender statistischer Hypothesen möglich. Darüberhinaus ist es wichtig zu unter-

entscheiden, ob es sich um eine eher explorative Fragestellung ('fundamental exploration', vgl. O'Neill & Wetherill, 1971) oder um eine möglichst strenge Prüfung wissenschaftlicher Hypothesen, die oft einseitige oder gerichtete statistische Hypothesen implizieren, handelt (vgl. Bredenkamp, 1980, S. 244; Hager & Westermann, 1983a, Kap. 8.3). Hager & Westermann argumentieren demzufolge für die durchgängige Verwendung von multiplen t-Tests sowohl bei der Prüfung von einzelnen oder mehreren einseitigen Hypothesen und von Mittelwertskontrasten wie auch bei der Prüfung von monotonen Trends. Hager & Meiler (1983) schlagen vor, alle Tests zu einer Familie zusammenzufassen, die der Prüfung einer wissenschaftlichen Hypothese dienen.

In den statistischen Studien über Teststärkeunterschiede von verschiedenen univariaten und multivariaten parametrischen multiplen Testverfahren ist die enge Beziehung zu den zu überprüfenden wissenschaftlichen Hypothesen nicht berücksichtigt worden. Die wichtigsten Ergebnisse der Arbeiten für das univariate c-Stichprobenproblem von Einot & Gabriel (1975), Welsch (1977) und besonders Ramsey (1978) wurden bereits in Abschnitt 5.4.1 angeführt. Für multivariate Verfahren hat Gabriel (1968, 1969b) wichtige Arbeiten publiziert (vgl. Abschnitt 5).

6.2 Bekannte Verfahren und Ergebnisse für Permutationstests

6.2.1 Asymptotische Teststärke

Aufbauend auf eine Arbeit von Lehmann & Stein (1949) für eine Verallgemeinerung des unverbundenen Zweistichprobenproblems hat Hoeffding (1952) nachgewiesen, daß unter sehr allgemeinen Bedingungen verschiedene univariate Permutationstests für ANOVA-Versuchspläne ((un-) verbundenes Zweistichprobenproblem, RB-k) für gegen unendlich strebenden Gesamtstichprobenumfang dieselbe asymptotische Teststärke gegen Lagealternativen besitzen wie die entsprechenden ANOVA-Teststatistiken mit Normalverteilungsannahme (vgl. auch Box & Andersen, 1955). Insbesondere für das Ein- und Zweistichprobenproblem sind auch neuere theoretische Arbeiten publiziert worden, die untersuchen, wie gut diese Annäherung für große endliche Stichprobenumfänge ist, d.h. was für eine sog. Konvergenzgeschwindigkeit sich ergibt (Albers, Bickel, & van Zwet, 1976; Robinson, 1973a für den RB-k). Falls die Normalverteilungsvoraussetzung erfüllt ist, besitzt der ubliche t-Test natürlich die gleichmäßig größte Teststärke.

Ohne ein festen Niveau α vorgeben zu müssen, können auch (asymptotische) Teststärke-Betrachtungen anhand des p-Wertes, des sog. beobachteten Signifikanzniveaus, angestellt werden. Da der p-Wert als Maß für die Evidenz gegen die Gültigkeit der Nullhypothese angesehen werden kann (vgl. z.B. Kempthorne, 1977), ist seine Große unter festen Alternativen ein interessantes Maß für die Teststärke.
Lambert & Hall (1982, 1983) betrachten u.a. den (bedingten) p-Wert für den Gesamtstichprobenumfang n

(6.1) $P_n(Y_n=y) = 1/|Q_n| \cdot |\{\pi \in Q_n | S_n(\pi y) \geq S_n(y)\}|$

für eine Teststatistik S_n, die bei Abweichungen von der Nullhypothese große Werte annimmt und für die Permutations-Untergruppe Q_n der symmetrischen Gruppe, unter der bei Gültigkeit der Nullhypothese die unbekannte Verteilung von Y_n permutationsinvariant ist. Sie weisen nach, daß unter Lagealternativen bei Normalverteilung und asymptotisch normalverteilten Teststatistiken,

(6.2) $-1/n \log P_n$

asymptotisch eine logarithmische Normalverteilung mit kompliziert berechenbarem Erwartungswert wie Varianz besitzt. Je größer dieser Erwartungswert und je kleiner die Varianz sind, umso größer ist die asymptotische Teststärke. Lambert & Hall (1982, Abschnitt 6) weisen nach, daß für das unverbundene Zweistichprobenproblem mit der Teststatistik (3.38) in der Form $S(Y)/n_1+n_2)$ asymptotisch übereinstimmung mit der Statistik (6.2) für den t-Test unter Normalverteilungsbedingungen besteht. Falls die wirkliche Alternative aber (nur leicht) von der hypothetischen abweicht, kann diese Optimalitätseigenschaft deutlich verletzt sein (vgl. Lambert, 1985 und Abschnitt 6.5.1).

Diese selbst für sehr einfache univariate Testprobleme mathematisch kompliziert zu beweisenden theoretischen Ergebnisse sind vornehmlich für eine theoretische Rechtfertigung der Permutationstests von Interesse. Von praktischem Nutzen sind dagegen eher Ergebnisse zur finiten Teststärke von Permutationstests.

6.2.2 Approximation der finiten Teststärke

Genauso wie für den asymptotischen Fall kann man versuchen, die Teststärke eines Permutations- oder Randomisierungstests für einen endlichen Stichprobenumfang zu approximieren. Die Vorgehensweise soll exemplarisch für das unverbundene Zweistichprobenproblem (vgl. Abschnitt 3.1.5 (i)) angedeutet werden. John & Robinson (1983b) betrachten das Randomisierungsmodell

$$
\begin{aligned}
Y_i &= U_{\pi(i)} + \theta \quad , \; i=1,..,n_1 \\
&\qquad\qquad \text{mit } \pi=\pi(1,..,n)\in S_n \\
Y_i &= U_{\pi(i)} \quad , \; i=n_1+1,..,n_1+n_2=:n
\end{aligned}
\tag{6.3}
$$

mit den unit values $U_{\pi(i)}$ und dem für alle Beobachtungseinheiten hypothetisch angenommenen, konstanten Effekt θ (Differenz der Effekte unter den beiden Bedingungen 1 und 2). Da die Verteilung der Teststatistik aus Abschnitt 3 auch bei einer festen Lagealternative nicht bekannt ist, da nicht alle Werte in der Teststatistik aus (3.36) gleichwahrscheinlich sind, muß man versuchen, für jeweils in einem Experiment beobachtete y_i die unbekannte Wahrscheinlichkeit

$$
P(S_\theta(y) \geq s^*_\alpha \mid u_1,..,u_n) \tag{6.4}
$$

bzw. die Teststärke des Zweistichproben-Randomisierungstests bei Gültigkeit der festen Lagealternative θ und den unit values $u_1,..,u_n$ zu approximieren. Dabei sei s^*_α der exakte kritische Wert zu einem erreichbaren α-Niveau (vgl. (2.3), wenn man den Randomisierungstest mit $y_1,..,y_n$ ausführt. Unter gewissen überprüfbaren (Regularitäts-) Bedingungen an die $u_1,..,u_n$ kann man eine Approximation an (6.4) für 'nicht zu weit von θ=0' entfernte Werte von θ angeben (vgl. John & Robinson, 1983b, Formel (2.13)).

Da diese Ergebnisse ebenfalls nur von theoretischem Interesse sind, wird der komplizierte mathematische Ausdruck hier nicht angegeben. Die approximative Teststärke ist von den unit values abhängig. Mit ähnlichen Methoden wie zur Approximation der Teststärke kann man auch den kritischen Wert s^*_α approximieren (vgl. Robinson, 1978, 1982). John & Robinson bemerken, daß Ergebnisse für das Ein- und das k-Stichprobenproblem analog zu gewinnen seien.

Die Ergebnisse zur Approximation der exakten Permutationsverteilung unter einer Lagealternative durch die entsprechende nichtzentrale F-Verteilung werden in Abschnitt 6.4 unter dem Thema Robustheitsstudien behandelt.

6.2.3 Finite konditionale und unkonditionale Teststärkebestimmung

Wie in Abschnitt 6.1.2 dargestellt, kann es für einen bedingten Test keine Bestimmung der Teststärke vor Beobachtung der Untersuchungsergebnisse geben, denn der kritische Wert des (Permutations-) Tests hängt von den tatsächlichen Beobachtungen ab. Ebenso ist die Verteilung der Teststatistik (bzw. der p-Wert) auch unter einer festen (Lage-) Alternative abhängig von den Beobachtungen. Am Beispiel des verbundenen Zweistichprobenproblems soll die Bestimmung der bedingten Teststärke erläutert werden (vgl. Gabriel & Hall, 1983; Gabriel & Hsu, 1983, Kempthorne, 1952; Kempthorne & Doerfler, 1969; Toothaker, 1972).

Wenn man festlegt, daß unter den beiden experimentellen Bedingungen n Beobachtungseinheiten betrachtet werden sollen, gibt es bekanntlich $|Q_n|=2^n$ verschiedene Möglich-

keiten, die beiden Bedingungen im Versuchsplan anzuordnen. Per Zufall wird eine dieser Anordnungen vom Experimentator ausgewählt; sei das D_h, $h \in \{1,..,|Q_n|\}$. Für sie bestimmt man etwa mit der Teststatistik (3.39) den zum Beobachtungsvektor y gehorenden p-Wert

$$p(y,D_h) = 1/|Q_n| \cdot |\{\pi \in Q_n | S(\pi y_{D_h}) \leq S(y_{D_h})\}| \quad .$$

Dann werden der Reihe nach alle anderen $|Q_n|-1$ möglichen Anordnungen von experimentellen Bedingungen betrachtet und angenommen, daß die tatsächlich unter D_n beobachteten Werte aus der experimentellen Anordnung D_l, $l \neq h$, $l \in \{1,..,|Q_n|\}$ resultieren. Anschließend wird jeweils für diese Anordnung der p-Wert

$$p(y,D_l) = 1/|Q_n| \cdot |\{\pi \in Q_n | S(\pi y_{D_l}) \geq S(y_{D_l})\}|$$

ermittelt. Die exakte bedingte Teststarke zum beobachteten Niveau $p(y,D_h)$ ist dann

(6.5)
$$q(y,p(y,D_h)) = 1/|Q_n| \cdot |\{D_l | p(y,D_l) \leq p(y,D_h)\}| \quad ;$$

d.h., die relative Häufigkeit, mit der die moglichen Experimente D_l, $l \in \{1,..,|Q_n|\}$ die Nullhypothese verwerfen würden zum Niveau $p(y,D_h)$.

Man kann statt des exakten p-Werts $p(y,D_l)$, $l=1,..,|Q_n|$, auch das exakte $(1-\alpha)$-Quantil für ein fur $|Q_n|$ erreichbares α-Niveau bestimmen: $s_{\alpha,l} := s_{|Q_n|(1-\alpha)}(y,D_l)$ und

(6.6)
$$q_{1-\alpha}(y,s_{\alpha,h}) = 1/|Q_n| \cdot |\{D_l | s_{\alpha,l} > s_{\alpha,h},\ l \neq h,\ l=1,..,|Q_n|\}|$$

die exakte bedingte Teststarke zum (erreichbaren) Niveau α ermitteln.

In mehrfaktoriellen Versuchsplänen ist, wie zuvor in den Abschnitten 3 und 4 dargestellt, zuerst eine (maximal-) invariante Transformation T=KY der Beobachtungen auszuführen und dann das zuvor dargestellte Verfahren auf den transformierten Beobachtungsvektor t anzuwenden.

Es ist leicht zu erkennen, daß die exakte bedingte Teststärke (6.5) nur für 'sehr kleine' Versuchspläne wirklich bestimmbar ist, denn es müssen $|\Omega_n|^2$ Berechnungen der Teststatistik erfolgen. Beispiele findet man in Gabriel & Hsu, 1983, Tabelle 1; Kempthorne & Doerfler, 1969, Tabelle 3 - 6. Für einen Stichprobenumfang von n=10 sind das für das verbundene Zweistichprobenproblem $(2^{10})^2 = 2^{20} = 1048576$ Permutationen, bzw. $2^{10} \cdot 2^{10-1}$ für einen zweiseitigen Test.

Als ein Ausweg wird empfohlen (vgl. z.B. Gabriel, 1979, Abschnitt 3.2), nur eine Teilmenge aller D_n, etwa eine Zufallsauswahl von $m << |\Omega_n|$ Plänen, zu betrachten und (6.5) damit zu schätzen (weiter vgl. Abschnitt 8.3).

Eine weitere Möglichkeit stellt die Bestimmung der sog. 'naiven' bedingten Teststärke dar (vgl. Gabriel, 1979; Gabriel & Hsu, 1983). Mit dieser versucht man, wie im parametrischen Fall vorzugehen, indem man den $(1-\alpha)$-Schwellenwert der exakten Permutationsverteilung der Teststatistik bei Gültigkeit der Nullhypothese bestimmt. Dieser sei mit $s_{u,1-\alpha}$ bezeichnet und $1-\alpha$ sei ein erreichbares Niveau. Da die unit values aber vor der Ausführung des Experiments nicht bekannt sind, kann man höchstens die aus dem Experiment geschätzten Effekte $\hat{y}$ von dem Beobachtungssektor y subtrahieren und für diese geschätzten unit values $\hat{u} = y - \hat{y}$ das exakte $(1-\alpha)$-Quantil $s_{\hat{u},1-\alpha}$ errechnen. Dazu bestimmt man $S(\pi\hat{u})$ für alle $\pi \in \Omega_n$ und wählt den $|\Omega_n| \cdot (1-\alpha)$ -ten geordneten Wert als Schwellenwert aus. Dann ermittelt man die

'exakte naive bedingte Teststarke zum Niveau α' mit Hilfe von:

(6.6) $q^*(y,s_{\hat{u},1-\alpha}) = 1/|\Omega_n| \cdot |\{\pi\in\Omega_n | S(\pi y) > s_{\hat{u},1-\alpha}\}|$.

Das bedeutet eine große Reduktion des Rechenaufwandes auf nur $2|\Omega_n|$ Permutationen. Gabriel & Hsu (1983) argumentieren, daß man bei dieser Vorgehensweise die Teststärke leicht überschätzt, da allgemein $s_{v_{D_h},1-\alpha} \geq s_{\hat{u},1-\alpha}$ ist. Für die von ihnen betrachteten Beispiele ist dieser Unterschied jedoch nicht sehr groß.

Die schon bei Kempthorne & Doerfler (1969) angewandte Methode vorauszusetzen, daß die unit values bekannte Werte annehmen, ist eigentlich nur von theoretischem Interesse. Die unit values selbst können nicht bekannt sein. Man kann sie nur aus den Beobachtungen des Experiments schätzen, indem man Schätzwerte (z.B. Kleinste-Quadratschätzer) für alle im linearen Modell vorhandenen Parameter von den Beobachtungen subtrahiert und diese Residuen als unit values ansieht. Für relativ kleine Stichprobenumfänge sind die Standardfehler der Schätzwerte aber sicher nicht unerheblich, sodaß es zu beträchtlichen Ungenauigkeiten in der Teststärkebestimmung kommen kann. Gabriel & Hsu argumentieren aber, daß gewisse Ungenauigkeiten in der Bestimmung der Teststärke zu tolerieren seien, denn sie schätzen die Sicherheit, mit der man eine Lagealternative exakt annehmen kann, auch nicht als perfekt ein. Weiterhin geben sie Formeln zur Schätzung der exakten Teststarke mit Hilfe der Normalverteilung an. Wie gut diese verschiedenen Ansätze im Vergleich ausfallen, ist bisher nicht untersucht worden.

Will man nicht die Teststärke bezüglich der im Experiment beobachteten Abweichung von der Nullhypothese analysieren, kann man nach Bestimmung des Schwellenwertes $s_{\hat{u},1-\alpha}$ aus (6.6) jeden beliebigen Effekt θ_o zu dem Residuenvektor $\hat{u}$ addieren und den Teststärke-Schätzwert q^* gegen diese Alternative θ_o nach Ausführung aller zulässigen Permutationen für ein festes α und einen festen Gesamtstichprobenumfang ermitteln.

In den ersten Studien zur Analyse der Teststärke von Permutationstests von Baker & Collier (1966b, 1968), Collier & Baker (1966) wurde ein anderer Zugang gewählt. Ausgehend von einer gegebenen Konstellation von unit values werden diese nach den im Modell zulässigen Permutationen vertauscht. Für jede Permutation werden die festen Effekte addiert und der Teststatistikwert errechnet. Für eine bestimmte Anzahl von zufälligen Vertauschungen der unit values wird dann die relative Häufigkeit bestimmt, mit der Teststatistikwerte größer als das entsprechende Quantil der Verteilung der Teststatistik unter Annahme der Normalverteilung (Quantil der zentralen F-Verteilung) ausfallen (vgl. auch Kempthorne & Doerfler, 1969, S. 239).

Verschiedentlich ist auch eine 'unkonditionale' Teststärke von Permutationstests verwendet worden. Toothaker (1972) und Still & White (1981) und Arnold (1964) sowie Royen (1978) für ein bivariates, verbundenes Zweistichprobenproblem generieren n-mal eine Residuen-Stichprobe aus einer bekannten Verteilung. Dann werden jeweils bestimmte Effektparameter addiert und ein exakter oder approximativer (vgl. Abschnitt 8) Permutationstest ausgeführt. Ein Schätzwert für die unkonditionale Teststärke ist dann die relative Häufigkeit von Ablehnungen der Nullhypothese bei vorgege-

benem Niveau α. Als Schwellenwert wird das zugehörige $(1-\alpha)$-Quantil der Verteilung der Teststatistik unter der Normalverteilungsannahme gewählt.

Allen bisher dargestellten Methoden der Teststärkebestimmung ist gemeinsam, daß jeweils nur die Teststärke bezüglich einer festen (Lage-) Alternative bestimmbar ist, sei sie die aktuell in den Beobachtungen vorliegende oder eine hypothetisch angenommene. In allen Fällen ist der Rechenaufwand beträchtlich bis unpraktikabel hoch (vgl. aber die Abschnitte 6.2.4 und 8.4.2). Für das univariate Ein- und Zweistichprobenproblem ist es aber möglich, den Permutationstest zu 'invertieren' und für alle möglichen Effektparameterwerte die Teststärke zu ermitteln. Das wird im folgenden Abschnitt erläutert.

6.2.4 Inversion von Permutationstests

Wie im vorhergehenden Abschnitt erläutert, ist die Bestimmung der exakten Teststärke für Permutationstests sehr mühsam oder praktisch undurchführbar. Als Ausweg wurden verschiedene approximative Möglichkeiten aus der Literatur diskutiert, die aber nur die (approximative) Teststärke für eine feste Alternative liefern.

Für das unverbundene univariate Zweistichprobenproblem (als Spezialfall eines allgemeineren Regressionsproblems) haben Gabriel & Hall (1983) ein Verfahren vorgestellt, mit dem mit demselben Aufwand wie für die Methoden aus Abschnitt 6.2.3 die gesamte (bedingte) Teststärke-Kurve für beliebige Lagealternativen ermittelt werden kann.

In einem ersten Schritt ist dazu die sog. Inversion des Permutationstests nötig. Man muß dazu das $(1-\alpha)$-Konfidenzintervall von Parameterwerten für den (Lage-) Parameter δ aus 3.1.5 (i) ermitteln, für das der p-Wert des Permutationstests gerade größer als α ausfällt. Für das Ein- und Zweistichprobenproblem sichern Monotonieeigenschaften bzgl. δ der Teststatistik bzw. des p-Wertes, daß sich ein Intervall von Parameterwerten ergibt, die mit Sicherheit $(1-\alpha)$ mit der Nullhypothese H $\delta = \Delta$ bzw. speziell $\delta = 0$ verträglich sind (vgl. Kempthorne & Doerfler, 1969; Lehmann, 1959, Kap. 5.12, Problem 23; Tritchler, 1984).

Getestet werden soll die Hypothese

H: $\delta = \Delta$, speziell $\Delta = 0$

(6.7) gegen

A: $\delta > \Delta$.

Als Statistik wählt man:

(6.8) $S_\Delta(\underline{Y}) = \bar{Y}_{1.} - \bar{Y}_{2.} - \Delta$.

Man kann zeigen, daß anstelle von S_Δ zur Bestimmung des p-Wertes $p^+(\Delta)$ für den Test von H gegen A:

(6.9) $p^+(\Delta) = 1/|S_n| \cdot |\{\pi \in S_n | S_\Delta(\pi\underline{y}) \geq S_\Delta(\underline{y})\}|$

- nur $K = \binom{n}{n_1}$ Permutationen aus S_n können zu unterschiedlichen Werten von S_Δ führen - folgende Statistik verwendet werden kann:

$$(6.10) \quad T_\pi^+(\underline{Y}) = \begin{cases} \bar{Y}_{1/2} - \bar{Y}_{2/1} & \pi \neq 1 \\ -\infty & \pi = 1 \end{cases} .$$

Mit $\bar{Y}_{1/2}$ sei das arithmetische Mittel aller Komponenten des Beobachtungsvektors $Y = (Y_{11},..,Y_{2n_2})'$ bezeichnet, die durch die Permutation π von der Stichprobe (Bedingung) 1 in die Stichprobe (Bedingung) 2 gelangen; $\bar{Y}_{2/1}$ ist analog definiert. Anstelle von (6.9) kann man äquivalent

(6.11) $p^+(\Delta) = 1/K \cdot |\{\pi \in \mathfrak{Q}_n | T_\pi^+(y) \geq \Delta\}|$

als (einseitigen) p-Wert zum Testen von H gegen A wählen; $\mathfrak{Q}_n$ bezeichnet dabei die Menge aller $K = \binom{n}{n_1}$ verschiedenen Vertauschungen von Beobachtungen zwischen den beiden Bedingungen. Nur diese können zu unterschiedlichen Werten $T_\pi^+(y)$ führen.

Mit (6.11) hat man eine Darstellung des p-Wertes als empirische Verteilungsfunktion der Statistik T_π^+ gefunden. Damit ist es möglich, mit nur K - 1 Berechnungen die p-Werte für jede Hypothese H: $\delta=\Delta$ und jede Alternative $\delta>\Delta$ aus der sog. p-Wert-Kurve abzulesen.

Analog definiert man zum Testen von H: $\delta=\Delta$ gegen A: $\delta<\Delta$:

(6.12) $p^-(\Delta) = 1/K \cdot |\{\pi \in \mathfrak{Q}_n | T^-(y) \geq \Delta\}|$

$$\text{mit } T_\pi^-(y) = \begin{cases} T_\pi^+(y) & \pi \neq 1 \\ +\infty & \pi = 1 \end{cases}$$

$p^-(\Delta)$ ist die (linksstetige) Überlebenskurve als Funktion von $T_\pi^-(y)$.

Zum Testen gegen die zweiseitige Alternative A: $\delta \neq \Delta$ kann man als p-Wert $2p^\pm(\Delta)$ mit

(6.13) $p^\pm(\Delta) = \min\{p^+(\Delta), p^-(\Delta), 1/2\}$

wählen (vgl. Abb. 6.1). Eine andere Methode zur Berechnung der p-Wert-Kurve stammt von Tritchler (1984, vgl. Abschnitt 8.3) und Hartigan (1969) sowie Forsythe & Hartigan (1970). p^+ und p^- sind also Treppenfunktionen in Abhängigkeit von Δ unter Einschluß von $\pm\infty$ mit Sprunghöhe $\alpha \cdot 1/K$ ($\alpha \in \mathbb{N}$) an den Stellen $\Delta = t_\pi^+(y)$ bzw. $t_\pi^-(y)$. Zum Testen für ein festes Niveau α und die Realisation y entscheidet man:

H_Δ^+ akzeptieren, falls $p^+(\Delta) \geq \alpha$

(6.14) H_Δ^- akzeptieren, falls $p^-(\Delta) \geq \alpha$

$H_\Delta^\pm$ akzeptieren, falls $p^\pm(\Delta) \geq \alpha/2$.

Ein- bzw. zweiseitige Konfidenzintervalle für die Realisation y sind ebenfalls aus der p-Wert-Kurve ablesbar mit Hilfe der sog. Quantilfunktion $Q(\alpha)$:

$Q^+(\alpha) = \inf\{\Delta | p^+(\Delta) > \alpha\}$ mit der unteren Konfidenzgrenze $Q^+(\alpha) \leq \Delta$ für ein einseitiges $(1-\alpha)$ - Konfidenzintervall für δ;

(6.15) $Q^-(\alpha) = \sup\{\Delta | p^-(\Delta) > \alpha\}$ mit der oberen Konfidenzgrenze $Q^-(\alpha) \geq \Delta$ für ein einseitiges $(1-\alpha)$ - Konfidenzintervall für δ;

$Q(\alpha_1) \leq \Delta \leq Q^-(\alpha_2)$ mit $\alpha_1+\alpha_2=\alpha$ als zweiseitigem $(1-\alpha)$-Konfidenzintervall für δ.

Ein sinnvoller (Punkt-)Schätzwert für δ ist dasjenige Argument bzw. die Intervallmitte von $p^\pm$, für das $p^\pm(\Delta)$ maximal wird:

(6.16) $\hat{\delta} = Q^+(1/2) = Q^-(1/2)$.

Alle p-Werte, Konfidenzintervalle und der Punktschätzer $\hat{\delta}$ können aus der p-Wert-Kurve abgelesen werden.

Beispiel: Sei $Y = (1, 3, 5, 0, 2)'$ mit $n_1 = 3$, $n_2 = 2$.

Getestet werden soll die Hypothese H: $\delta=\Delta= 1$ gegen A: $\delta > 1$. Es gibt $K=\binom{5}{3}=10$ zu betrachtende Permutationen. Die Werte von $S_\Delta(y)$ bzw. $T_\pi^+(y)$ sind in der folgenden Tabelle 6.1 angegeben; die zugehörige $p^\pm$-Kurve befindet sich in Abb. 6.1.

Permu-tation	Stichprobe 1	Stichprobe 2	Teststatistik $S_{\Delta=1}(\pi y)$	Rang	$T_{\pi}^{\pm}(y)$
1	1,3,5	0,2	3- 1 - 1 = 1	(9)	$\bar{+}\ \infty$
2	1,3,0*	5,2	4/3-7/2 - 1 =-19/6	(2)	5-0 = 5
3	1,3,2	0,5	2-5/2 - 1 =- 3/2	(4.5)	5-2 = 3
4	1,0,5	3,2	2-5/2 - 1 =- 3/2	(4.5)	3-0 = 3
5	1,2,5	0,3	8/3-3/2 - 1 = 1/6	(7.5)	3-2 = 1
6	0,3,5	1,2	8/3-3/2 - 1 = 1/6	(7.5)	1-0 = 1
7	2,3,5	0,1	10/3-1/2 - 1 = 11/6	(10)	1-2 =-1
8	0,2,5	1,3	7/3- 2 - 1 =- 2/3	(6)	2-1 = 1
9	0,3,2	1,5	5/3- 3 - 1 =- 7/3	(3)	3-1 = 2
10	1,0,2	3,5	1 - 4 - 1 =- 4	(1)	4-1 = 3

*Unterstrichene Werte sind durch die Permutation π zwischen den beiden Stichproben ausgetauscht worden

Tab. 6.1.: Beispiel: Berechnung der Teststatistikwerte $T_{\pi}^{\pm}(y)$ zur Inversion des unverbundenen Zweistichproben-problems

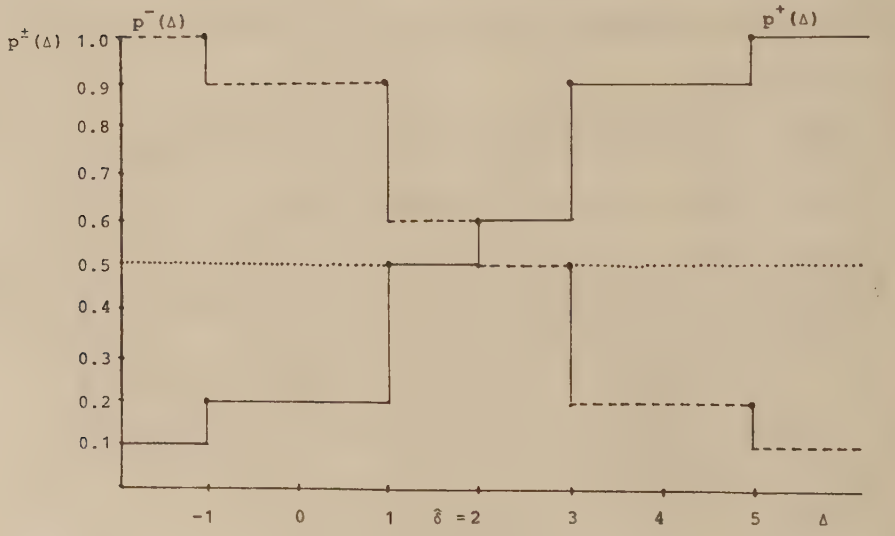


Abb. 6.1.: p-Wert Kurve für die Beispieldaten aus Tabelle 6.1; relevant ist der Kurvenanteil unterhalb von $p^{\pm}(\Delta)=0.5$

Der p^+-Wert für die obige Hypothese ist $p^+(\Delta=1)=2/10$. $Q^+(0.4)=1$, d.h. mit 60% Sicherheit ist $\Delta>1$.

Die Statistiken $T_\pi^\pm(y)$ lassen sich aber auch benutzen, um die exakte konditionale Teststärke $\beta(\delta)=(\delta,\Delta,\alpha,\hat{u})$ mit $K(K-1)$ Berechnungen für alle Hypothesen Δ gegen alle Lage-Alternativen δ zu bestimmen. Mit der Methode aus Abschnitt 6.2.3 waren K^2 Berechnungen für eine einzige der unendlich vielen (Δ,δ)-Kombinationen erforderlich. Subtrahiert man den geschätzten oder vermuteten Effekt $\hat{\Delta}$ von den Beobachtungen y, d.h. $\hat{u}=y-\hat{\Delta}$, hat man:

$$T_\pi(\hat{u}) = T_\pi(y) - \Delta$$

bzw. allgemein für beliebige Lage-Verschiebungen

(6.17) $$T_\pi(y) = T_\pi(\hat{u}) + \delta \quad .$$

Die empirische Verteilungsfunktion von $T_\pi(y)$ mit dem angenommenen Effekt δ erhält man somit durch eine einfache Verschiebung der empirischen Verteilungsfunktion für die (geschätzten) unit values.

Entsprechend werden die Quantilfunktionen für z.B. H_Δ^+ zum Niveau α als das Komplement des Annahmebereiches gegeben durch:

$$\Delta < Q^+(y,\alpha) \quad .$$

Nach (6.18) hat man, falls die Beobachtungen y unter der Alternative δ entstanden sind:

$$\Delta - \delta < Q^+(\hat{u},\alpha) \quad .$$

Damit ist die exakte konditionale Teststärke im unverbundenen Zweistichprobenproblem für einen Niveau-α Test von H_Δ^+ bei Vorliegen der Alternative δ und unit values $\hat{u}$ gegeben durch:

(6.19) $$\beta^+(\delta,\Delta,\alpha,\hat{u}) = 1/K! \cdot |\{\pi\in Q_n | Q^+(\pi\hat{u},\alpha) > \Delta-\delta\}|$$
$$= 1/K! \cdot |\{\pi\in Q_n | -Q^+(\pi\hat{u},\alpha) < \delta-\Delta\}| \quad .$$

Das ist aber nichts anderes als die (linksstetige) empirische Verteilungsfunktion der negativen Quantilfunktionswerte in Abhängigkeit von $\delta-\Delta$. Äquivalent gilt:

(6.20) $\beta^{+}(\delta,\Delta,\alpha,\hat{u}) = 1/K \cdot |\{\pi \in \underline{Q}_n | -t_{\pi,(l)}(\pi\hat{u}) < \delta-\Delta\}|$

mit $(l)=[K\cdot\alpha]$ der größten ganzen Zahl kleiner oder gleich $K\alpha$,

denn $Q^{+}(\alpha)=t_{\pi,(l)}(y)$ für $l/K<\alpha<(l+1)/K$ (vgl. Abb. 6.1)

Man muß also für jede mögliche der $K=\binom{n}{n_1}$ Anordnungen π_ν, $\nu=1,..,K$ von (geschätzten) unit values die $K-1$ von $-\infty$ verschiedenen Werten von $T^{+}_{\pi_\nu}(\hat{u})$ berechnen und danach den $l=[K\alpha]$-kleinsten geordneten Wert multipliziert mit -1 $-t_{\pi_\nu,(l)}(\hat{u})$ bestimmen.

Die empirische Verteilungsfunktion dieser K α-Quantile $-t_{\pi_\nu,(l)}(\hat{u})$ stellt die Teststärke-Kurve für alle Hypothesen H^{+}_{Δ} gegen alle einseitigen Alternativen $\delta>\Delta$ dar.

Beispiel: Für die Beobachtungsdaten aus Tabelle 6.1 sei ein geschätzter Effekt (bzw. eine Effektdifferenz von $\hat{\Delta}=2$ für die Stichprobe 1 angenommen. Als Vektor der unit values hat man damit $\hat{u}=(-1,1,3,0,3)'$.

Für jede der $K = 10$ möglichen Zuweisungen der unit values zu den beiden experimentellen Bedingungen sind in jeder Spalte der Tabelle 6.2 die Werte der $T^{+}_{\pi_\nu}(\pi\hat{u})$ - Statistik für die jeweils $K - 1 = 9$ übrigen Anordnungen angegeben.

Permutation	Teststatistik $t^{+}_{\pi}(\pi\hat{u})$										
	1	2	3	4	5	6	7	8	9	10 (=k)	
1	-	-3	-1	-1	1	1	3	1	0	-1	
2	3	-	2	2	2	4	3	2.5	3	1	
3	1	-2	-	0	2	1	4	1.5	1	-3	
4	1	-2	0	-	2	2	3	3	0.5	-1	
5	-1	-2	-2	-2	-	0	2	1	-0.5	-3	
6	-1	-4	-1	-2	0	-	2	1	-1	-1.5	
7	-3	-3	-4	-3	-2	-2	-	-1	-3	-2.5	
8	-1	-2.5	-1.5	-3	-1	-1	1	-	-2	-4	
9	0	-3	-1	-0.5	0.5	1	3	2	-	-4	
10	1	-1	3	1	3	1.5	2.5	4	4	-	
$t_{\pi,(1)}(\pi\hat{u})$	-3	-4	-4	-3	-2	-2	1	1	-3	-4	$\alpha = 1/10$
	3	4	4	3	2	2	-1	-1	3	4	
$t_{\pi,(2)}(\pi\hat{u})$	-1	-3	-2	-3	-1	-1	2	1	-2	-4	$\alpha = 2/10$
	1	3	2	3	1	1	-2	-1	2	4	
$t_{\pi,(3)}(\pi\hat{u})$	-1	-3	-1.5	-2	0	0	2	1	-1	-3	$\alpha = 3/10$
	1	3	1.5	2	0	0	-2	-1	1	3	

Tab. 6.2.: Exakte konditionale Teststärke für die Beispieldaten aus Tabelle 6.1.

Zur Berechnung aller Tabelleneinträge sind nur die K(K-1)/2 Werte unter der Diagonalen zu ermitteln, da die Tabelle stets schiefsymmetrisch ist. Für jede Spalte kann man dann bestimmte α-Quantile (sinnvollerweise für erreichbare α - Niveaus) nach Ordnen der Statistikwerte angeben. Dieses sind die $t^{+}_{\pi_\nu}(\pi\hat{u})$ -Werte, bei deren Auftreten als Ergebnis des Experimentes ein Test zum Niveau α ein signifikantes Ergebnis anzeigen würde. Die (linksstetige) empirische Verteilungsfunktion dieser Quantile für jeweils ein ausgewähltes α ist die gesuchte exakte Teststärkefunktion in Abhängigkeit von der Differenz aus echtem festem Effektparameter δ und dessen Wert Δ bei Gültigkeit der Nullhypothese. Die Teststärke-Kurve ist in Abbildung 6.2 dargestellt.

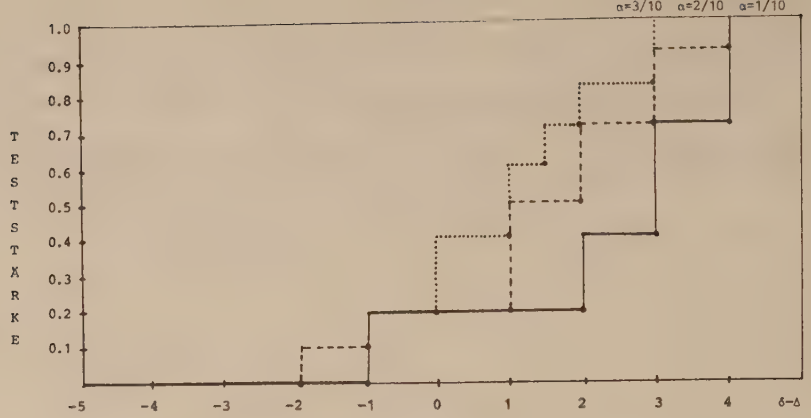


Abb. 6.2.: Exakte konditionale Teststärke des Tests $H:\delta \leq \Delta$ gegen $\delta > \Delta$ für die Beispieldaten aus Tabelle 6.1 und für die α-Niveaus 0.1, 0.2, 0.3. Die Teststärke für $\delta - \Delta = 0$ und $\alpha = 2/10$ ist identisch mit dem empirischen p-Wert des Tests (vgl. Abb. 6.1).

Die Kurve gibt gerade die Anzahl von Anordnungen der unit values an, die bei einem bestimmten Effekt $\delta - \Delta$ zur Verwerfung der Nullhypothese $\delta = \Delta$ führen. Das ist aber die Definition der exakten konditionalen Teststärke. Für $\delta - \Delta = 1$ ergibt sich das exakte Niveau des Tests $p^+ = 0.2$, wie man schon aus der Spalte mit den Statistikwerten von $S_\Delta(\pi y)$ in Tabelle 6.1 entnehmen konnte.

Analog kann man auch die Teststärke für H_Δ^- gegen $A: \delta < \Delta$ bestimmen:

$$(6.21)\quad \beta^-(\delta,\Delta,\alpha,\hat{u}) = 1/K \cdot |\{\pi \in \Omega_n \mid Q^-(\pi\hat{u},\alpha) < \Delta - \delta\}|$$
$$= 1/K \cdot |\{\pi \in \Omega_n \mid t^-_{\pi,(K-m)}(\pi\hat{u}) < \Delta - \delta\}|$$
und $(K-m)$ so, daß $[(m+1)/K] \geq \alpha$,

denn $Q^-(\alpha) = t_{(K-m)}(y)$ für $(m+1)/K \geq \alpha > m/K$ (vgl. Abb. 6.1).

Für große Werte von K ist die Generierung der Tabelle 6.2 allerdings ebenfalls sehr aufwendig oder nicht praktikabel. Eine Approximation der exakten Teststärke ist jedoch leicht möglich, indem man nur eine (Zufalls-) Stichprobe aus der

Menge aller möglichen Anordnungen der unit values benutzt (vgl. Abschnitt 8.3). Genauso ist es auch möglich, durch Ziehen von Zufallsstichproben von unit values aus einer spezifizierten Verteilung die unkonditionale Teststärke zu approximieren.

Für die approximativen Verfahren bleibt aber weiterhin gültig, daß man für alle möglichen festen Lagealternativen δ-Δ in einem Rechengang die gesamte approximative Teststärkekurve ermitteln kann.

Diese vorteilhafte Situation ist aber auf Testprobleme beschränkt, in denen eine sog. 'pivotal statistic' (Ankerstatistik, vgl. Gabriel & Hall, 1983) existiert, in der die Verteilung unter jeder (Lage-) Alternative durch eine einfache Verschiebung, die durch den unbekannten (Lage-) Parameter determiniert ist, aus der bekannten Verteilung unter der Nullhypothese hervorgeht (vgl. (6.17) oder (6.18)).

Das zuvor dargestellte Verfahren läßt sich auch auf das verbundene Zweistichprobenproblem mit derselben Hypothese wie (6.7) sowie den Test auf Symmetrie um Δ gegen eine Lageverschiebung im Einstichprobenproblem anwenden. Als Statistik kann man in diesem Fall mit (3.39)

(6.22) $S_\Delta(Y) = Y_{1.}-Y_{2.}-n\Delta$

verwenden. Ω_n ist jetzt die Gruppe aller $K=2^n$ Intra-Block-permutationen. Als Ankerstatistik ist

(6.23) $$T_\pi^{\pm}(Y) = \begin{cases} Y_{1/2}-Y_{2/1} & \pi \neq 1 \\ \mp\infty & \pi = 1 \end{cases}$$

geeignet.

Eine Verallgemeinerung des unverbundenen Zweistichprobenproblems auf den Fall, daß zusätzlich eine Schichtungsvariable (z.B. organismische Variable) mit $j=1,..,k$ Stufen

vorliegt, ist ebenfalls naheliegend.

Dann wählt man mit $Y_j=(Y_{j11},\ldots,Y_{j1n_{j1}},Y_{j21},\ldots,Y_{j2n_{j2}})'$

$Y=(Y_1',\ldots,Y_k')'$ und $n_j=n_{j1}+n_{j2}$; $n=\sum_{j=1}^{k} n_j$:

(6.24) $S_\Delta^j(Y_j)=\bar{Y}_{j1.}-\bar{Y}_{j2.}-\Delta$ und $S_\Delta(Y)=\sum_{j=1}^{k} S_\Delta^j(Y_j)$

als Ankerstatistik:

(6.25) $$T_\pi^{\pm}(Y)=\begin{cases}\sum_{j=1}^{k}(\bar{Y}_{j1/2}-\bar{Y}_{j2/1}) & \pi\neq 1\\ \mp\infty & \pi=1\end{cases}$$

$\bar{Y}_{j1/2}$ bezeichnet den Mittelwert der durch die Permutation π von Bedingung 1 nach Bedingung 2 permutierten Beobachtungen in Schicht j. Zulässige Vertauschungen sind alle $K=\prod_{j=1}^{k}\binom{n_j}{n_{j1}}$ Intraschichtpermutationen.

6.3 (Approximative) Bestimmung der finiten konditionalen und unkonditionalen Teststärke

6.3.1 Exakte und geschätzte konditionale Teststärke für Permutations- und Randomisierungstests

Die Überlegungen aus Abschnitt 6.2.3 zum Randomisierungstest für das univariate Zweistichprobenproblem sollen hier auf alle uni- und multivariaten linearen Modelle, für die Invarianz-Permutationstests existieren, übertragen werden. Zur Vereinfachung der Darstellung wird der univariate Fall ausführlich behandelt. Die multivariate Verallgemeinerung ist stets in offensichtlicher Analogie zu leisten.

Wie in (3.1) - (3.3) sei ein verteilungsfreies lineares Modell

$$Y = \theta + \varepsilon,\ Y \sim F_\theta(y) = F(y-\theta)\ ,\ F(x) = F(\pi x)$$
$$\text{für alle } \pi \in \mathcal{Q}_n^*;\ \theta \in U \subset R^n,\ x, y \in R^n$$

gegeben.

Die zu testende lineare Hypothese ist (vgl. (3.6) und (4.13)): H: $K\theta = 0$ gegen A: $\delta = K\theta \neq 0$.

In Multiparametermodellen mußte u.U. eine (maximal-) invariante Transformation $KY = K\theta + K\varepsilon$ mit $\mathcal{Q}_n$-invarianter Verteilung des transformierten Fehlervektors $K\varepsilon$ ausgeführt werden, um die 'störenden' Parameter zu eliminieren (vgl. die Abschnitte 3.1.3 bzw. 4.3).

Bei der Wahl einer geeigneten Teststatistik S, die bei Annahme von 'großen' Werten gegen die Nullhypothese spricht, gilt, daß unter H alle Realisationen von $S(\pi KY)$ unter zulässigen Permutationen $\pi \in \mathcal{Q}_n$ dieselbe Wahrscheinlichkeit $1/|\mathcal{Q}_n|$ besitzen. Zur (nichtrandomisierten) Testentscheidung überprüft man für die Realisation y von Y, ob

$$p^{\geq}(Ky) = 1/|\mathcal{Q}_n| \cdot |\{\pi \in \mathcal{Q}_n \mid S(\pi Ky) \geq S(Ky)\}|$$

ein vorgegebenes α-Niveau überschreitet oder nicht (vgl. 3.18)).

Wie in der parametrischen ANOVA und MANOVA analysiert man naheliegenderweise die Teststärke gegenüber Lagealternativen, d.h.

$$\delta = K\theta \neq 0 \quad .$$

Nur bei Gültigkeit der Nullhypothese sind alle $S(\pi KY)$, $\pi \in \mathcal{Q}_n$, gleichwahrscheinlich. Unter der Alternative dagegen ist die Verteilung von S unbekannt. Man kann sich aber leicht überlegen, daß bei Gültigkeit der Lagealternative gilt:

(6.26) KY und alle $Z_{\pi,\delta} = \delta + \pi(KY - \delta)$, $\pi \in \mathcal{Q}_n$
sind identisch verteilt.

Nach Verschiebung von KY um δ bleiben lediglich die K_ε übrig, deren Verteilung, auch wenn sie um δ verschoben wird, invariant ist unter den bei Gültigkeit der Nullhypothese zulässigen Permutationen $\pi \in \mathcal{Q}_n$ (vgl. Pyhel, 1980b, Abschnitt 6). Alle Realisationen von $Z_{\pi,\delta}$ unter dem Permutationsmodell haben also ebenfalls dieselbe (bedingte) Wahrscheinlichkeit $1/|\mathcal{Q}_n|$.

Sei deshalb für s = S(Ky)

(6.27)
$$q_s^{>}(Ky) = 1/|\mathcal{Q}_n| \cdot |\{\pi \in \mathcal{Q}_n | S(\delta + \pi(Ky - \delta) > s\}|$$

die Anzahl von zulässigen Permutationen, die nach Verschiebung in die Alternative δ zu Teststatistikwerten führen, die größer als die Teststatistik für den transformierten Beobachtungsvektor selbst ausfallen. Man kann beweisen, daß dann gilt:

(6.28)
$$E_{F_\delta^{KY}}(q_s^{>}(KY)) = 1 - F_\delta^S(s) \quad ,$$

d.h., die (bedingte) Erwartung von (6.27) bezüglich der transformierten Beobachtungen KY, verteilt nach der unbekannten Verteilungsfunktion F_δ unter der festen Alternative δ ist gleich 1 minus der unbekannten Verteilungsfunktion F_δ^S der Teststatistik S(KY) unter der Alternative δ berechnet für die Realisation s = S(KY). Anstelle von $q_s^{>}$ kann man auch für ein mit dem Permutationstest erreichbares Niveau α (d.h. $|\mathcal{Q}_n| \cdot \alpha \in \mathbb{N}$) definieren:

(6.29)
$$q_\alpha^{>}(Ky) = 1/|\mathcal{Q}_n| \cdot |\{\pi \in \mathcal{Q}_n | S(\delta + \pi(Ky - \delta) > s_{1-\alpha}\}|$$

mit $s_{1-\alpha}$ als kritischem Wert des Tests zum Niveau α.

Die Funktion $q_s^{>}$ bzw. $q_\alpha^{>}$ ist eine erwartungstreue Schätzfunktion für die unbekannte Wahrscheinlichkeit, bei Gültigkeit der Alternative δ einen Wert $S(Ky) > s$ bzw. $> s_{1-\alpha}$ zu

beobachten. Deshalb wird $q_s^>$ bzw. $q_\alpha^>$ als die geschätzte konditionale Teststärke zum beobachteten Niveau $p^>(Ky)$ bzw. zum festen Niveau α bezeichnet.

Da die Verteilung der Teststatistik unter der Alternative δ nicht bekannt ist, ist nur eine (Punkt-)Schätzung der bedingten Teststärke möglich. Für ihre praktische Bestimmung muß vom transformierten Beobachtungsvektor Ky die feste hypothetische Lagealternative δ subtrahiert werden. Dieser 'Residuen'-Vektor $Ky-\delta$ wird der Reihe nach allen unter der Nullhypothese zulässigen Permutationen unterworfen, δ wird addiert und man überprüft, ob der beobachtete Teststatistikwert, der zu einem bestimmten p-Wert geführt hatte, überschritten wird. Statt des beobachteten Statistikwertes s kann auch der Vergleich mit dem exakten Schwellenwert $s_{1-\alpha}$ zum erreichbaren Niveau α erfolgen. In beiden Fällen sind $2|\Omega_n|$ Berechnungen eines Teststatistikwertes erforderlich, um neben dem p-Wert den Teststärkewert (6.27) bzw. (6.29) für einen festen Wert δ zu bestimmen.

Wenn man zu einem Randomisierungsmodell übergeht, ist die exakte konditionale Textstärke zu erreichbarem Niveau α und einer festen Alternative δ zumindest theoretisch bestimmbar. Man nimmt jede der $|\Omega_n|$ zulässigen Permutationen π_h der Komponenten des transformierten Beobachtungsvektors als Realisation eines neuen Experimentes. Dann führt man für jeden dieser $|\Omega_n|$ Versuchspläne $D_h, h=1,..,|\Omega_n|$ einen exakten Permutationstest unter Ausführung aller $|\Omega_n|$ zulässigen Vertauschungen der transformierten Beobachtungen aus und bestimmt die Anzahl von Fällen, in denen der Test zu einem signifikanten Ergebnis geführt hat, d.h.

(6.30) $q_\alpha^{>}(Ky) = 1/|\Omega_n| \cdot |\{D_h | S(\pi_h Ky) > s^{D_h}_{|\Omega_n|(1-\alpha)}, \pi_h \in \Omega_n, h=1,..,|\Omega_n|\}|$

Äquivalent kann man, ohne den jeweiligen exakten Schwellenwert zum erreichbaren Niveau α bestimmen zu müssen, betrachten:

(6.31) $q_\alpha^{>}(Ky) = 1/|\Omega_n| \cdot |\{D_h | p^{\geq}(\pi_h Ky) \leq \alpha, \pi_h \in \Omega_n, h=1,..,|\Omega_n|\}|$.

Durch die konsekutive Wahl aller möglichen unter der Nullhypothese zulässigen Vertauschungen der transformierten Beobachtungen als (hypothetisch) beobachtbare Werte unter jedem der möglichen Versuchspläne erhält man die finite Verteilung der $p^{\geq}$-Wert-Statistik für ein festes α, einen festen Stichprobenumfang mit der transformierten Realisation Ky und für eine feste unbekannte Alternative. Die Größe $q_\alpha^{>}(Ky)$ kann somit sinnvoll interpretiert werden (vgl. Kempthorne & Doerfler, 1969).

Zur Bestimmung von $q_\alpha^{>}(Ky)$ sind jedoch $|\Omega_n|^2$ Berechnungen des Teststatistikwertes erforderlich. Deshalb ist in den meisten Fällen nur eine Schätzung von (6.31) möglich. Man wählt systematisch oder zufällig eine Teilmenge von Versuchsplänen aus, etwa vom Umfang 1000 (vgl. Gabriel & Hall, 1983, Abschnitt 11) und betrachtet auch nur eine Zufallsstichprobe aller zulässigen Permutationen (vgl. Abschnitt 8.3).

Über die Güte des Schätzwertes $\hat{q}_\alpha^{>}(Ky)$ gibt es keine Untersuchungen. Gabriel & Hsu (1983, Abschnitt 6) vertreten aber die Meinung, daß jede Teststarkebestimmung notwendigerweise nur eine grobe Abschätzung sein kann, da die wahre Alternative nicht vor Ausführung des Experiments bekannt sein kann.

6.3.2 'Naive' Bestimmung der finiten konditionalen Teststärke

Die im vorhergehenden Abschnitt dargestellten Methoden zur Bestimmung der konditionalen Teststärke, wie sie in (6.27) bzw. (6.29) definiert ist, sind in der Regel praktisch nicht durchführbar, da δ nicht bekannt ist bzw. die Anzahl $|\mathcal{Q}_n|^2$ durchzuführender Berechnungen astronomisch hoch ausfällt. Als Ausweg ist folgende Vorgehensweise möglich, die von Gabriel & Hsu (1983) wegen der Analogie zum Vorgehen im parametrischen Fall als 'naiv' bezeichnet wird.
Nach Abschnitt 3.1.3 gibt die (orthogonale) Projektion K_A in den zum Hypothesenunterraum orthogonalen Raum die 'Abweichung' von der Nullhypothese wieder. Damit ist δ schätzbar durch:

$$\hat{\delta} = K_A y = K_A(Ky) \quad \text{(vgl. (3.16))}.$$

Als 'Residuenvektor' kann man dann definieren:

(6.32) $$\hat{u} = Ky - K_A y = (K-K_A)y \quad .$$

Im einfaktoriellen Blockplan gilt beispielsweise (vgl. Abschnitt 3.1.4):

$$K-K_A = I_{rk} - 1/k I_r \otimes J_k - 1/r J_r \otimes I_k + 1/rk J_{rk}$$

$$= (I_r - 1/r J_r) \otimes (I_k - 1/k J_k)$$

und angewendet auf die (i,j)-te Komponente von y:

$$\hat{u}_{ij} = ((K-K_A)y)_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \quad i=1,..,r,\ j=1,..,k \quad .$$

Damit sind die aus der ANOVA bekannten Kleinsten-Quadratschätzer $\hat{\alpha}_i, \hat{\beta}_j, \hat{\mu}$ von y_{ij} subtrahiert worden. Jede Komponente des Residuenvektors erhält man für alle Versuchspläne aus Kapitel 3 durch Differenzbildung aus dem Mittel über alle Indizes von KY, die nicht auch als Indizes in dem auf

Null zu testenden Parameter vorkommen. Beispielsweise hat man im Split-Plot Plan für den Test von H_o: $(\alpha\beta)_{ij} = 0$ für alle i=1,..,r; j=1,..,k (vgl. (3.6.1)) als Komponenten des Residuenvektors (vgl. Abschnitt 3.1.5(xi)):

$$\hat{u}_{ihj} = (Ky)_{ihj} - (\bar{\bar{K}}y)_{i.j} = y_{ihj} - \bar{y}_{ih.} - \bar{y}_{i.j} + \bar{y}_{i..} \quad .$$

Man bestimmt dann in Analogie zu (6.27):

(6.33) $$\hat{q}_s^{>}(\delta,\hat{u},Ky) = 1/|\Omega_n| \cdot |\{\pi\in\Omega_n | S(\hat{\delta}+\pi\hat{u}) > s\}| \quad .$$

Die Analogie zum parametrischen Fall liegt darin, daß man den kritischen Wert s aus der 'Nullverteilung', d.h. den unter Annahme der Gültigkeit der Nullhypothese bestimmten exakten, zum beobachteten Signifikanzniveau $p^{\geq}(Ky)$ gehörenden Schwellenwert benutzt. Zur Bestimmung der geschätzten Teststärke zu einem erreichbaren festen Niveau α kann man dann ähnlich wie in (6.29) mit dem $[|\Omega_n|(1-\alpha)]$-Quantil $s_{1-\alpha}$ berechnen:

(6.34) $$\hat{q}_\alpha^{>}(\hat{\delta},\hat{u},Ky) = 1/|\Omega_n| \cdot |\{\pi\in\Omega_n | S(\hat{\delta}+\pi\hat{u}) > s_{1-\alpha}\}| \quad .$$

Dazu sind $2|\Omega_n|$ Berechnungen von Werten der Teststatistik notwendig.

Es ist natürlich auch möglich, ein bestimmtes $\delta = \delta_o$ fest vorzugeben und (6.33) bzw. (6.34) mit δ_o anstelle von $\hat{\delta}$ zu ermitteln.

Zur 'naiven' Approximation von (6.29) müßte man die Permutationsverteilung von $S(\hat{u})$ unter allen bei Gültigkeit der Nullhypothese zulässigen Permutationen $\pi\in\Omega_n$ und daraus insbesondere das erreichbare $(1-\alpha)$-Quantil $s_{1-\alpha}(\hat{u})$ ermitteln. Die naive Teststärkeschätzung ist dann:

(6.35) $$\hat{q}_\alpha^{>}(\hat{\delta},\hat{u},Ky) = 1/|\Omega_n| \cdot |\{\pi\in\Omega_n | S(\pi Ky) > s_{1-\alpha}(u)\}| \quad .$$

Auch in diesem Fall müssen $2|\mathfrak{Q}_n|$ Teststatistikwerte berechnet werden. Für große $|\mathfrak{Q}_n|$ ist wieder zur praktischen Bestimmung der naiven Teststärke eine (Zufalls-) Stichprobe aller möglichen zulässigen Permutationen auszuwählen.

6.3.3 Approximative apriori-Bestimmung der finiten, unkonditionalen Teststärke von Permutationstests

Zur Schätzung der bedingten Teststärke nach Ausführung eines Experiments sind die zuvor dargestellten Verfahren geeignet. Man erhält mit (6.34) bzw. (6.35) einen Schätzwert für die Teststärke bei bestimmtem Niveau α, oder bei dem aktuellen Niveau $p^{\geq}$, bei fest vorgegebenem Gesamtstichprobenumfang n für die aktuell erhobenen Beobachtungen mit einer aus diesen Beobachtungen geschätzten Abweichung $\hat{\delta} \neq 0$ von der Nullhypothese. Diesen Schätzwert für die bedingte Teststärke sollte man demnach bei Ausführung eines Permutationstests mitberechnen, auch wenn die Anzahl auszuführender Permutationen mit anschließender Berechnung der Teststärke sich gegenüber dem ausschließlichen Testen der Nullhypothese verdoppelt. Der Vektor $\hat{\delta}$ kann dann leicht zu einer Schätzung des Zählers in den verschiedenen Definitionen eines Nichtzentralitätsparameters verwendet werden (vgl. Hager & Westermann, 1983a, Kap. 9.3). Für den einfaktoriellen Blockplan hat man z.B.:

$$(6.36) \qquad \hat{\lambda}^2 = |K_A y|^2/\hat{\sigma}^2_{Fehler} = r\sum_{j=1}^{k}(\bar{y}_{.j}-\bar{y}_{..})^2/\hat{\sigma}^2_{Fehler} \quad \text{(vgl. (3.21))} .$$

Einen Schätzwert für die Fehlervarianz erhält man aus

$$(6.37) \qquad \hat{\sigma}^2_{Fehler} = |(K-K_A)y|^2/rg(K-K_A) = |\hat{u}|^2/rg(K-K_A) .$$

Für den einfaktoriellen Blockplan ergibt das:

$$\hat{\sigma}^2_{Fehler} = \sum_{i=1}^{n} \sum_{j=1}^{k} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 / (r-1)(k-1) \quad .$$

Die Planung eines Experimentes, insbesondere die Abschätzung eines minimal notwendigen Gesamtstichprobenumfangs n^*, mit dem ein Mindesteffekt spezifizierter Größe bei Einhaltung von α- und β-Fehler aufgedeckt werden kann, ist dagegen so nicht möglich. Da im parametrischen Fall die Verteilung der Teststatistik unter der Lagealternative bekannt ist, sind dort Vertafelungen möglich (vgl. Hager & Möller, 1983), aus denen n^* leicht berechenbar ist. Für das Pillai'sche Spurkriterium ist auch im multivariaten Fall diese Berechnung möglich (vgl. Bredenkamp, 1980; Bredenkamp & Erdfelder, 1985; Hager & Möller, 1983).

Man sieht hier deutlich, was für starke Modellannahmen in der ANOVA und MANOVA notwendig sind, um die apriori-Festlegung des notwendigen Stichprobenumfanges zu erreichen. Nur wenn man σ^2_{Fehler} bzw. die Fehler-Varianz-Kovarianz-Matrix Σ_{Fehler} für den multivariaten Fall als bekannt voraussetzt, ist für uni- und multivariate Permutationstests folgendes Vorgehen denkbar:

Da für die asymptotische relative Effizienz (ARE) vom Permutations- zum ANOVA-Test gegen eine Lagealternative unter Normalverteilungsannahme ARE = 1 gilt (vgl. Lehmann, 1959 und Abschnitt 6.1.1, sowie die Argumentation von Bredenkamp (1980, S. 74) bei Verwendung von Rangtests sowie Guenther, 1982), legt man für ein fixiertes, mit dem Permutationstest erreichbares α sowie β dieselben Tafeln von Cohen (1977) bzw. Hager & Möller (1983) wie im parametrischen Fall zugrunde. Im Gegensatz zur (M)ANOVA muß aus

dem vollständig spezifizierten Effektvektor δ_o der erwarteten Abweichungen von der Nullhypothese der Effektparameter λ^2 oder f^2 (vgl. z.B. Bredenkamp, 1980, S. 54; Hager & Westermann, 1983a, S. 160) unter Verwendung von σ^2_{Fehler} zuerst bestimmt werden. Dann kann man den mit oben erwähnten Tabellen minimal notwendigen Gesamtstichprobenumfang n^* ablesen.

Mit einem Simulationsexperiment (vgl. Abschnitt 8.2) kann man überprüfen, ob die angestrebte Teststärke $1-\beta$ in einem echten Experiment tatsächlich erreichbar sein kann. Dazu generiert man n_o-mal Zufallsstichproben von Residuenvektoren u_ν, $\nu = 1,..,n_o$ der Dimension n^* mit $\hat{\sigma}^2_{Fehler} \approx \sigma^2_{Fehler}$, addiert jeweils den erwarteten Effektvektor δ_o, $y_\nu = \delta_o + u_\nu$, und führt mit einer Zufallsstichprobe aller möglichen, unter der jeweiligen Nullhypothese zulässigen Permutationen n_o-mal einen Permutationstest durch (vgl. Taylor & Thompson, 1986). Als Schätzwert der exakten Teststärke betrachtet man unter Verwendung von (6.30) bzw. (6.31) und der Teststatistik $S(Ky_\nu)$:

(6.38) $\hat{q}^{>}_{\alpha}(\delta_o) = 1/n_o \cdot |\{\nu=1,..,n_o | p^{\geq}(Ky_\nu) \leq \alpha\}|$, $y_\nu \in R^n$.

Wenn $\hat{q}^{>}_{\alpha}(\delta_o) \approx 1-\beta$ ist, kann das reale Experiment ausgeführt werden.

Nur unter Vorgabe von δ_o, ohne Kenntnis von σ^2_{Fehler} kann man die Bestimmung von (6.38) für eine 'realistische' Wahl des Gesamtstichprobenumfanges mit in 'sinnvollen' Schritten vergrößertem/verkleinertem Umfang solange wiederholen, bis eine geschätzte Teststärke von ungefähr $1-\beta$ resultiert (Jöckel, 1979). Die Verteilung, aus der Zufallsstichproben von Residuen gezogen werden, sollte 'so gut wie möglich' den Kenntnisstand über die für die spezielle experimentelle

Fragestellung in Vorexperimenten gefundenen empirischen Verteilungen widerspiegeln. Wie z.B. Guenther (1982) für das Ein-, Zwei- und k-Stichprobenproblem in Simulationsstudien demonstriert, sind je nach zugrundeliegenden Verteilungen die minimal nötigen Gesamtumfänge dennoch sehr ähnlich zum Fall der Normalverteilung.

Die in Abschnitt 6.2.4 vorgestellten Methoden zur Generierung der gesamten Teststärke-Kurve für das Zweistichprobenproblem sind zur Bestimmung des notwendigen n^* ebenfalls ungeeignet. Man kann nur für bestimmte aus der parametrischen Situation plausible n nach Berechnung oder Schätzung der Teststärkekurve ablesen, ob der erwartete Effekt mit hinreichender Teststärke feststellbar ist. Dazu müssen aber auch wieder Informationen über mögliche (nichtnormale) Residuenverteilungen vorliegen.

6.3.4 (Approximative) Bestimmung der Teststärke von multiplen Testprozeduren

Die bisher in Abschnitt 6.3 beschriebenen Vorgehensweisen können prinzipiell auch zur Analyse der Teststärke multipler Testprozeduren zum multiplen Niveau α verwendet werden.

(i) Wenn es sich um multiple Tests für die d Komponenten multivariater Beobachtungen handelt (vgl. Abschnitt 5.3), läßt sich die (naive) konditionale Teststärke zum erreichbaren multiplen Niveau α in Analogie zu (6.27) bzw. (6.33) und (6.34) schätzen. Man bestimmt den geschätzten Effektvektor $\hat{\delta}=(\hat{\delta}_1',..,\hat{\delta}_d')'$, der jetzt verglichen mit (6.33)

von d-facher Länge ist und den Residuenvektor $\hat{u}$. Für jede unter der globalen Nullhypothese $H_o = \bigcap_{i=1}^{d} H_o^i$ zulässige Vektor-Permutation aus der Gruppe $\hat{\mathfrak{G}}_n = \mathfrak{G}_n \otimes \{1_d\}$ werden die multiplen p-Werte $p^{[i]}$, $i=1,..,d$, wie in Abschnitt 5.3.1 dargestellt, ermittelt. Wenn man sich nur für die Elementarhypothesen interessiert, kann man folgenden, geschätzten d-dimensionalen Teststärke-Vektor definieren:

(6.39) $\hat{q}_\alpha^{>}(\hat{\delta},\hat{u},\hat{\underline{Ky}}) = (\hat{q}_{1,\alpha}^{>},..,\hat{q}_{d,\alpha}^{>})'$ mit

$\hat{q}_{i,\alpha}^{>}(\hat{\delta}_i,\hat{u}_i,(\hat{\underline{Ky}})_i)$

als relative Häufigkeit von Permutationen aus $\hat{\mathfrak{G}}_n$, für die in der sequentiellen multiplen Testprozedur zum multiplen Niveau α der zugehörige multiple p-Wert $p^{[i_\nu]}$ - $i_\nu \in \{1,..,d\}$- zu einer Ablehnung der Elementarhypothese H_o^i geführt hat.

Statt $\hat{\delta}$ könnte (6.39) für jeden hypothetisch angenommenen Effekt δ_o bestimmt werden.

(ii) Genauso ist in Verallgemeinerung von (6.30) die exakte konditionale Teststärke bestimmbar:

(6.40) $\underline{q}_\alpha^{>}(\hat{\underline{Ky}}) = (\underline{q}_{1,\alpha}^{>},..,\underline{q}_{d,\alpha}^{>})'$

mit $\underline{q}_{i,\alpha}^{>}$, $i=1,..,d$, als der relativen Häufigkeit von Versuchsplänen D_h, $h=1,..,|\hat{\mathfrak{G}}_N|$, für die in der sequentiellen multiplen Testprozedur zum multiplen Niveau α der zugehörige multiple p-Wert $p^{[i_\nu]}$, $i_\nu \in \{1,..,d\}$, zu einer Ablehnung der Elementarhypothese H_o^i, d.h. $p^{[i_\nu]} \leq \alpha$ geführt hat.

Wie zuvor, ist nur bei 'doppelter' Zufallsauswahl von Versuchsplänen und je Plan von zulässigen Permutationen die multiple Teststärke (6.40) für die feste unbekannte Alternative schätzbar.

(iii) Im Fall von multiplen Tests zum Vergleich von k Faktorstufen ist die Analyse der Teststärke der Paarvergleiche je zweier Faktorstufen von besonderem Interesse. Ramsey (1978) unterscheidet mehrere Fälle:

- all-pairs power: Wahrscheinlichkeit, alle wahren Unterschiede zwischen Paaren von Faktorstufen zu entdecken;
- any-pair power: Wahrscheinlichkeit, mindestens einen wahren Unterschied zwischen Faktorstufenpaaren zu entdecken;
- per-pair power: Wahrscheinlichkeit, wahre Unterschiede für die einzelnen Paare zu entdecken.

Alle drei Arten von Teststärke sind auch auf Teilmengenhypothesen mit mehr als zwei Faktorstufen verallgemeinerbar. Für die Bestimmung oder Schätzung der bedingten Teststärke ist nur die per-pair- oder allgemeiner per-subset-Teststärke von Interesse. Da nicht bekannt sein kann, welche Paare sich tatsächlich unterscheiden, ist die all-pairs power nicht ermittelbar. Die any-pair Teststärke dürfte wie im parametrischen Fall (vgl. Ramsey, 1978, S. 482) ebenfalls nicht interessant sein, da sie ähnlich hoch wie die globale Teststärke bzgl. der globalen Nullhypothese ausfallen dürfte.

Bei allen in Abschnitt 5.4.1 behandelten Verfahren von Ryan, Peritz oder Welsch wie auch bei der Bonferroni- oder der Bonferroni-Holm Prozedur ist für die Paarhypothesen darauf zu achten, ob bei Verwerfen einer Paarhypothese für permutierte Beobachtungen die Relation der Faktorstufenmittel dieselbe ist wie für die unpermutierten Beobachtungen. Bei Kenntnis der wahren Unterschiede würde man sonst einen sog. Fehler dritter Art begehen. Dieser ist stets nicht größer als $\alpha/2$, wenn α als das Niveau des Paarvergleichs angesetzt ist. In Analogie zu (6.31) kann man also je Paarvergleich

zwischen den Stufen $i,j \in \{1,..,k\}$ zwei Teststärkewerte unterscheiden:

(6.41)
$q^{>}_{i<j,\alpha}$, falls die sequentielle multiple Testprozedur die zweiseitige $H_o^{i,j}$ verwirft und $\hat{\mu}_{i,D_h} < \hat{\mu}_{j,D_h}$;
$q^{>}_{i>j,\alpha}$, falls die sequentielle multiple Testprozedur die zweiseitige $H_o^{i,j}$ verwirft und $\hat{\mu}_{i,D_h} > \hat{\mu}_{j,D_h}$.

Die exakte konditionale Teststärke zum multiplen Niveau α für die gesamte multiple Testprozedur wird dann durch die beiden $k(k-1)/2$-komponentigen Vektoren

(6.42)
$$\underline{q}^{>}_{<,\alpha} = (q^{>}_{1<2,\alpha}, q^{>}_{1<3,\alpha}, \cdots, q^{>}_{(k-1)<k,\alpha})'$$

$$\underline{q}^{>}_{>,\alpha} = (q^{>}_{1>2,\alpha}, q^{>}_{1>3,\alpha}, \cdots, q^{>}_{(k-1)>k,\alpha})'$$

wiedergegeben. Wie zuvor ist (6.42) nur bei doppelter Zufallsauswahl von Versuchsplänen D_h und jeweils unter $H_o^{i,j}$ zulässigen Permutationen realistischerweise schätzbar.

Die naive Schätzung der konditionalen Teststärke ist analog prinzipiell nur wie in (6.34) durchführbar. Dort müssen allerdings das Quantil $s_{1-\alpha}$ durch das entsprechende Quantil zum Niveau der verwendeten multiplen Testprozedur ersetzt und gegebenenfalls Implikationsrelationen wie in der Peritz-Prozedur beachtet werden.

Wenn jedoch Erwartungen darüber bestehen, daß nur eine der beiden Richtungen $i>j$ oder $i<j$ gültig ist, kann eine $k(k-1)/2$-elementige Kombination von Komponenten der beiden Vektoren aus (6.42) betrachtet werden. Dann ist aber fraglich, ob nicht eine andere Strategie, in der gerichtete Hypothesen z.B. mit multiplen Permutationstests für das (un-) verbundene Zweistichprobenproblem überprüft werden, teststärker ist (vgl. Abschnitt 7.1).

Für Vergleiche gegen eine Kontrollbedingung (vgl. Abschnitt 5.4.2) kann man analog i.a. mit einseitigen Elementarhypothesen vorgehen.

6.3.5 Diskussion

Wie schon erläutert, ist das Problem der Optimalität von Permutationstests, insbesondere deren Teststarkeeigenschaften, bisher wenig untersucht. Ein interessanter, mathematisch jedoch sehr aufwendiger und anspruchsvoller Ansatz sind die Approximationen von asymptotischer und finiter Teststärke für das (un) verbundene Zweistichprobenproblem. Es zeigt sich, wie für einen bedingten Test nicht anders zu erwarten, die Abhängigkeit von den sog. unit values. Die Abhängigkeit von den tatsächlich im Experiment beobachteten Werten verhindert die Vertafelung der Teststatistikwerte selbst unter einer festen Lagealternative. Das ist bekanntlich für die Verteilung jeder Permutationsteststatistik unter Gültigkeit der Nullhypothese auch der Fall. Die Stärke der Modellannahmen im parametrischen (M)ANOVA-Modell, für das bei festen Effekten eine nichtzentrale F-Verteilung resultiert, wird dadurch deutlich.
Eine apriori-Bestimmung der konditionalen (!) Teststärke selbst für ein spezielles Experiment ist ebenfalls unmöglich. Man kann nur - wie dargestellt - mehr oder weniger grobe Schätzungen bei vollständiger, explizierter Spezifizierung aller Komponenten des Lageparametervektors unter der speziellen Lagealternative vornehmen, wenn man für die Residuen eine bestimmte Verteilung annimmt und aus dieser eine Zufallsstichprobe zieht, oder wenn man aus den Residuen

vorhergegangener Experimente mit identischen oder ähnlichen unabhängigen und abhängigen Variablen Stichproben bildet. Eine verläßliche Residuenstichprobe setzt aber sicher eine umfangreiche Erforschung einer wissenschaftlichen Fragestellung voraus. Gute Schätzungen der bedingten (aposteriori) Teststärke sind dagegen mit den vorgestellten Verfahren mit einigermaßen vertretbarem Rechenaufwand zu gewinnen. Diese aposteriori-Teststärke sollte bei Ausführung eines Permutationstests neben der Angabe des Schätzwertes für den gesamten Effektvektor über die Bestimmung des p-Wertes hinaus mitgeteilt werden. Inwieweit naive Teststärke und geschätzte exakte konditionale Teststärke übereinstimmen, ist bisher für komplexere Versuchspläne nicht untersucht. Auch für multiple Testprozeduren sind beide letzteren konditionalen Teststärkeschatzungen prinzipiell ausführbar. Der Rechenaufwand ist i.a. jedoch beträchtlich. Zudem ist bei multiplen Testprozeduren bezüglich Faktorstufen das Problem zweiseitiger Alternativen zu beachten bzw. das Problem von Fehlern dritter Art bei Festlegung einseitiger Alternativen. Für diese ist fur Permutationstests bisher nicht untersucht, ob nicht die Anwendung einseitiger Permutationstests für (un-)verbundene Zweistichprobenprobleme teststarker ausfällt.

6.4 Robustheit von uni- und multivariaten Randomisierungs- und Permutationstests

6.4.1 Definitionen von Robustheit

In Anlehnung an Box (1953) wird ein statistisches Verfahren allgemein als 'robust' bezeichnet (vgl. z.B. auch Kanji, 1975), wenn es wenig empfindlich gegen Verletzungen von Annahmen ist, die in dem Modell, auf dem das Verfahren basiert, enthalten sind. Diese 'Definition' ist jedoch nicht spezifisch genug. Oft ist dabei nur an die Einhaltung des nominellen α-Niveaus beim Testen einer statistischen Hypothese gedacht worden (vgl. Bradley, 1968, S. 25). Tatsächlich wird in verschiedenen Robustheitsstudien für ANOVA-Tests auch untersucht, ob der Fehler 2. Art bzw. die Teststärke, sich ebenfalls nicht (wesentlich) verändert (Baker & Collier, 1966b, 1968; Collier & Baker, 1966; Still & White, 1981). Scheffe (1959, S. 361) hat bereits darauf aufmerksam gemacht, daß die annahernde Einhaltung der Teststärke unter Modellverletzungen beim Testen gegen eine bestimmte Alternative ebenfalls wichtig ist.

In einer vehementen Kritik vieler Studien zur Robustheit varianzanalytischer Verfahren führt Bradley (1978) an, daß diese häufig zu positiv eingeschätzt wird. Er fordert zudem eine Quantifizierung des Robustheitsbegriffs derart, daß maximal tolerierbare relative Abweichungen zwischen nominellem und tatsächlichem α-Niveau zu spezifizieren sind. Er schlägt ein 'strenges' Kriterium mit maximalen absoluten Abweichungen von $\alpha/10$ und ein 'liberales' Kriterium mit $\alpha/2$-Abweichungen vor. Ähnliche Forderungen könnten für die Teststärke aufgestellt werden. Hager, Lübbeke & Hübner (1983)

sowie Hager & Westermann (1983a) fordern explizit, die Robustheit bezuglich α - und β -Fehler beim Testen einer bestimmten Hypothese zu analysieren.

Eine Übersicht über weitere, mathematisch kompliziertere Konzepte von Robustheit eines statistischen Tests gibt Lambert (1982). Für Permutationstests ist ihr Konzept der 'qualitativen Robustheit' von Interesse. Dazu wird untersucht, wie sich der p-Wert beim Testen einer Hypothese mit der Teststatistik S (vgl. Abschnitt 2), genauer wie sich $-1/n \log p_S(X)$, mit X n-dimensionaler Zufallsvektor der n Beobachtungen - verändert, wenn X oder die Verteilung von X sich ändern. Ein Test ist dann - anschaulich gesprochen - qualitativ robust, wenn kleine Änderungen von X oder der Verteilung von X mit kleinen Änderungen von $-1/n \log p_S(X)$ einhergehen. Lambert (1982) beweist, daß die Permutationstests für das Ein- und das unverbundene Zweistichprobenproblem nicht qualitativ robust sind, genauso wie die entsprechenden t-Tests. In Lambert (1985) sind weitere neuere Untersuchungen zitiert, die zeigen, daß die Teststärke des Zweistichproben-Permutationstests für bestimmte Verteilungen unter der Alternative stark abnimmt. Mit diesen Befunden stimmt eine Bemerkung von Green (1981) in der Besprechung des Buches von Edgington überein. In eigenen nicht publizierten Robustheitsstudien hat er gezeigt, daß der t- und sein analoger Permutationstest unter denselben Abweichungen - viele Ausreißer, ungewöhnliche Verteilungen - dasselbe Ausmaß an Nichtrobustheit zeigen.

Lambert (1985) schlagt deshalb vor, eine robuste Teststatistik für den (Zweistichproben-) Permutationstest zu wählen. Extreme Beobachtungen werden zensiert (vgl. Hager &

Westermann, 1983a, S. 117f für eine Literaturübersicht) und durch denselben Zensierungswert ersetzt. Dann wird der Permutationstest wie üblich ausgeführt (vgl. Abschnitt 3.1.5 (i)). Andere Vorschläge zur Wahl einer robusten Teststatistik sind die Ersetzung des arithmetischen Mittels (Nair, 1940; Randles & Wolfe, 1979, S. 348, 354) oder die Verwendung von sog. Hodges-Lehmann Schätzern (Randles & Wolfe, 1979, S. 348) oder sog. M-Schätzern (Huber, 1964, Andrews et al., 1972; Wainer, 1976; Schrader & McKean, 1977; Randles & Wolfe, 1979) oder anderen robusten Lageschätzern (Lee & Fung, 1985).

6.4.2 Univariate Randomisierungs- und Permutationstests

Wie Scheffe (1959, S. 323ff) darstellt, wird der übliche F-Test der ANOVA mit Normalverteilungsannahme, in dem man den beobachteten Wert der Teststatistik mit den vertafelten Werten der zentralen F-Verteilung vergleicht, als gute Approximation an einen exakten Randomisierungs- oder Permutationstest mit derselben Teststatistik angesehen (vgl. auch Bredenkamp, 1980, S. 21; Ogawa, 1974; Steinhorst, 1982; Wehrhahn & Ogawa, 1982). Es ist auch vorgeschlagen worden, die Teststärke des Randomisierungstests mithilfe der nichtzentralen F-Verteilung zu approximieren (Matts & McHugh, 1983).

In der ANOVA wird für jeden n-dimensionalen Beobachtungsvektor $Y = y$ die Prüfgröße $F(y)$ mit dem vom speziellen y unabhängigen Schwellenwert $F_{1-\alpha,\nu_1,\nu_2}$ verglichen. Für den exakten Permutationstest ist der kritische Wert das erreichbare $(1-\alpha)$-Quantil der Permutationsverteilung $F_{(1-\alpha)|Q_n}(y)$,

das von der Realisation y abhängt. Für verschiedene Versuchspläne ist nachgewiesen worden, daß bei Multiplikation der Freiheitsgrade ν_1, ν_2 mit einem von y abhängigen Korrekturfaktor ϕ der exakt nur aufwendig zu bestimmende Schwellenwert gut zu approximieren ist. Für den einfaktoriellen Blockplan z.B. hat man (vgl. Scheffe, 1959, S. 321f):

(6.43) $\phi = 1/(1-V/r)-2/(r(k-1))$,

wobei V ein Maß für die Variabilität der Blockvarianzen ist (für den CR-k siehe Box & Andersen, 1955, S. 13; für das Lateinische Quadrat LS-k, Welch, 1937, S. 41).

Neben asymptotischen Überlegungen (z.B. Wald & Wolfowitz, 1944) ist die 'gute Übereinstimmung' zwischen ANOVA- und Permutations-F-Test auf drei Arten demonstriert worden:

(i) Für bestimmte tatsächlich beobachtete y ergab sich gute Übereinstimmung (z.B. Pitman, 1937a; Welch, 1937; Kempthorne, 1952).

(ii) Statt der F-Statistik wird die Beta-verteilte Teststatistik $U = SS_h/(SS_h + SS_e)$ betrachtet (vgl. (4.15) für den MANOVA-Fall und Ito, 1980, S. 219f) und die ersten zwei bis vier Momente von U unter den üblichen ANOVA-Voraussetzungen mit denen unter dem Permutationsmodell verglichen (z.B. RB-K: Scheffe, 1959, S. 324; CR-K: Pitman, 1938, S. 333ff; partiell balanzierter Blockplan: Rao, 1963, S. 35; LS-k: Welch, 1937).

(iii) Mit Fortschritten in der Computer-Technologie war es möglich, in mehr oder weniger umfangreichen Monte-Carlo Studien (vgl. Abschnitt 8.2) für artifizielle Daten ohne (Collier & Baker, 1963, 1966) und mit spezifizierbaren Abweichungen von den ANOVA-Voraussetzungen wie Schiefe und

abweichende Wölbung der unit values im einfaktoriellen Zufallsplan (Baker & Collier, 1966b), Block-Behandlungs-Wechselwirkungen und/oder Heterogenitat der Blockvarianzen im RB-k-Plan (Baker & Collier, 1968) und deren Kombination für den Split-Plot Plan (Collier, Baker & Mandeville, 1967) für den Fall der Nullhypothese und verschieden starke Abweichungen von ihr die Übereinstimmung in α-Niveau und Teststärke zwischen ANOVA-F-Test und analogem Permutationstest zu analysieren. Obwohl die Anzahl betrachteter Faktorstufen, Blockanzahlen, α-Niveaus, β-Niveaus, Effektmuster und unit-values Vektoren recht gering ist, wird eine gute Übereinstimmung sowohl für Niveau und Teststärke über einen weiten Schiefe- und Wölbungsbereich, sowie reduziertes Niveau und reduzierte Teststärke für kombiniert auftretende Block-Behandlungs-Interaktionen und Blockvarianzen-Heterogenität berichtet.

Problematisch an den Studien von Baker & Collier (1966b, 1968) und Collier und Baker (1966) ist, daß sie beim Vergleich von ANOVA- und Permutations-F-Test mit einem 'fixed cut-off point' (Collier & Baker, 1966, S. 199; Collier, Baker & Mandeville, 1967) arbeiten. Es wird zur Bestimmung von p-Wert und 'naiver' Teststärke des Permutationstests $F(\pi y)$ stets mit dem festen α- bzw. $1-\beta$ Schwellenwert der zentralen bzw. nichtzentralen F-Verteilung verglichen, wobei man eine recht kleine Zufallsauswahl von nicht mehr als 1000 unter der Nullhypothese zulässigen Permutationen für ausreichend hält. Nur in der Arbeit von Collier & Baker (1966) wurden die exakten Quantile der approximativen Permutationsverteilung zum Vergleich herangezogen. Kempthorne et al. (1961) finden für das verbundene Zweistichprobenproblem mit verschiedenen Datensätzen aus verschiedenen Verteilungen

für n = 6, 8, 10 z.T. ein erhöhtes Niveau für den Fall, daß der fixed cut-off point verwendet wurde. Toothaker (1971) benutzt für den RB-3 Versuchsplan mit 5 Blöcken und 1000 zufälligen Permutationen für verschiedene Grade von Blockvarianzheterogenität und Block-Behandlungs-Interaktionen zwei Modifikationen des festen F-Quantils als Schwellenwert. Der Korrekturfaktor ϕ aus (6.43) wird entweder einmal für die Originaldaten, oder das arithmetische Mittel $\bar{\phi}$ über alle 1000 zufällig ausgewählten Permutationen der Daten gebildet. Wegen der verschiedenen Block-Behandlungs-Interaktionseffekte verändert sich ϕ ja unter Permutation der Daten. Zusätzlich wird zum ersten Mal in einer größeren Robustheitsstudie auch der p-Wert des Permutationstests bestimmt. Mit $\bar{\phi}$ erzielt man die beste Übereinstimmung mit dem exakten Niveau.

Toothaker (1972) analysiert in einer weiteren größeren Robustheitsstudie für das unverbundene Zweistichprobenproblem mit normalverteilten, gleichverteilten oder χ^2_3 - verteilten Beobachtungen das Niveau und die unkonditionale Teststärke des t-Tests und des Permutationstests mit der t-Statistik als Teststatistik. Die Übereinstimmung zwischen beiden Verfahren war sowohl für das Niveau wie für die Teststärke hoch. Für die betrachtete schiefe Verteilung war die Teststärke des Permutationstests besonders dann größer als die des t-Tests, wenn bei ungleichem Stichprobenumfang die kleinere Stichprobe den größeren Mittelwert aufwies. In den meisten Fällen war der Unterschied allerdings sehr gering. Zu einem ähnlichen Ergebnis kommt Murphy (1967) für ein nominelles α-Niveau von 5% und eine unkonditionale Teststärkeberechnung für je 500 - 750 Replikationen auch dann,

wenn bei ungleichen Varianzen die beiden Stichprobenumfänge identisch sind. Erwähnt sei noch einmal die Bemerkung von Green (1981) aufgrund eigener nicht publizierter Robustheitsstudien, daß Permutationstests und t-Tests in sehr ähnlichem Maße anfällig gegen Voraussetzungsverletzungen seien.

In der umfangreichen Studie von Still & White (1981) für $\alpha = .05$ bzw. .01, sieben abgestufte Abweichungen von der Nullhypothese und je 1000 Replikationen für die Normalverteilung und eine schiefe, zweigipflige Verteilung, gibt es für CR-3 und RB-3 gute Übereinstimmung zwischen F- und Permutationstest und eine Überlegenheit für den Permutationstest bei der nicht normalen Verteilung.

Edgington & Ezinga (1978) zeigen für verschiedene Beispiele von Zweistichprobenproblemen, daß einzelne Ausreißerwerte den p-Wert des Permutationstests weniger stark vergrößern als den des t-Tests. Derselbe Effekt wird für das verbundene Zweistichprobenproblem berichtet (Edgington, 1980, S. 127f).

6.4.3 Multivariate Randomisierungs- und Permutationstests

Außer der Vermutung von Kempthorne (1977) wie auch Edgington (1980), daß multivariate Permutationstests aufgrund ihrer strukturell nichtlinearen Teststatistiken möglicherweise weniger robust gegen Verletzungen von Modellannahmen seien, gibt es nur wenige Arbeiten, die das Problem (empirisch) untersuchen. Auch für das parametrische Problem wird die Situation in den Sammelreferaten von Ito (1969, 1980) nur wenig anders eingeschätzt. Dort wird zusammenfassend für den mehrdimensionalen einfaktoriellen Zufalls-

plan angegeben, daß für mehrere der üblichen Teststatistiken ähnliche Robustheitseigenschaften wie für den ANOVA-Fall vorliegen, daß diese Übereinstimmung jedoch mit steigender Dimensionenzahl abnimmt.

Bisher sind lediglich einfaktorielle Probleme betrachtet worden. Von Ito (1980) wird nur die theoretische Arbeit von Mardia (1971) für den pCR-k-Plan erwähnt. Mardia wählt den Zugang (ii) aus Abschnitt 6.4.2, indem er die ersten beiden Momente der Permutationsverteilung des Pillai'schen Spurkriteriums berechnet und diese durch die ersten beiden Momente einer Beta-Verteilung ohne und mit adjustierten Freiheitsgraden approximiert. Für zweidimensionale Beobachtungen, $k = 2$, $\alpha = 5\%$ und $n_1 + n_2 = 10$ ergeben sich für verschiedene von der zweidimensionalen Normalverteilung verschiedene Verteilungen und je 1000 Replikationen für die fixed cut-off point Methode (vgl. Abschnitt 6.4.2) mit adjustierten Freiheitsgraden sehr dicht am nominellen α-Niveau gelegene p-Werte. Für nicht adjustierte Freiheitsgrade konnte die Abweichung absolut $|\alpha-p| = .0211$ betragen.

Arnold (1964) hat für das zweidimensionale verbundene Zweistichprobenproblem mit Hotelling's T^2-Statistik für drei Verteilungstypen, $n = 4$, 6, 8 und $\alpha = .05, .025, .01, .005$ für weniger als 1000 Replikationen ebenfalls mit adjustierten Freiheitsgraden der Beta-Verteilung anhand der beiden ersten Momente der Permutationsverteilung gute Übereinstimmung mit dem nominellen Niveau zumindest für das liberale Kriterium von Bradley erzielt.

6.4.4 Diskussion

In noch stärkerem Maße als für die Teststärke gilt, daß die Robustheitseigenschaften von Permutationstests bisher als zweitrangiges Problem behandelt worden sind. Hauptziel sowohl für ANOVA- wie in geringerem Umfang für MANOVA-Probleme war es nachzuweisen, daß die Teststatistiken mit Normalverteilungsvoraussetzungen ähnliche Eigenschaften wie unter dem Permutationsmodell ohne diese Voraussetzung besitzen. Weiterhin war von Bedeutung, daß bei Verletzung weiterer (M)ANOVA-Modellvoraussetzungen wie Varianzheterogenität oder Block-Behandlungs-Interaktionen (M)ANOVA-Teststatistiken und dieselben Teststatistiken unter dem Permutationsmodell ähnliche Veränderungen in bezug auf Niveau und Teststärke zeigen. Diese Ergebnisse sind aber nicht völlig einheitlich. Die Abhängigkeit von der Anzahl gezogener Stichproben aus einer bestimmten Verteilung zur Schätzung von p-Wert und Teststärke, von der Anzahl und Art betrachteter Residuenverteilungen, vom Stichprobenumfang je Faktorstufe(nkombination), von der Anzahl Faktorstufen je Faktor, von der Anzahl abhängiger Variablen in multivariaten Plänen, von Ausmaß und Art der Verletzung von Block-Behandlungs-Interaktionen, vom Ausmaß an Heterogenität der Varianz bzw. Varianz-Kovarianz Matrix, von Ausmaß und Muster additiver Effekte und von der Güte der (Pseudo-) Zufallszahlengeneratoren zur Erzeugung der Residuenstichproben machen die Untersuchungen von Robustheitseigenschaften zu einem sehr aufwendigen Unternehmen. Aus diesem Grund werden in dieser Arbeit keine eigenen Untersuchungen zur Robustheit von uni- und besonders multivariaten Permutationstests vorgelegt.

Wie für die Analyse der Teststarke gibt es keine Robustheitsstudien, in denen Modellverletzungen bei Residuenverteilungen, die im Permutationsmodell zulassig sind, betrachtet werden (vgl. die Bemerkungen von Scheffe (1959, Fußnoten S. 323, S. 325)). Wenn man Permutationstests als eigenständige, praktisch einsetzbare Testverfahren akzeptiert, ist diese Fragestellung wichtiger als der Vergleich mit dem Normalverteilungsfall. Das gilt insbesondere auch für die MRPP-Statistiken (vgl. Abschnitt 4.7). Speziell die Güte der p-Wert Approximation dürfte nicht unbeeinflußt sein von solchen Abweichungen. Über die Robustheit multipler Testverfahren unter einem Permutationsmodell liegen ebenfalls keine Arbeiten vor. Auch die Verwendbarkeit robuster Statistiken (vgl. z.B. Schrader & McKean, 1977), d.h. robuster Schätzer anstelle von Kleinste-Quadratschätzern in den Summen quadrierter Abweichungen, wurde noch nicht überprüft.

7. BEDEUTUNG VON RANDOMISIERUNGS- UND PERMUTATIONSTESTS FÜR DIE UNI- UND MULTIVARIATE DATENANALYSE IN DER PSYCHOLOGIE

7.1. Bedeutung für eine deduktivistische Theorie des Experimentierens

Neben der Variierbarkeit von Bedingungen (sog. unabhangige Variablen, Faktoren) durch den Versuchsleiter ist die Kontrolle störender Einflüsse durch zufällige Zuteilung von Beobachtungseinheiten (experimentelle Randomisierung) zu den Bedingungen die zweite definierende Eigenschaft eines (psychologischen) Experimentes (vgl. Bredenkamp, 1980; Campbell & Stanley, 1966). Bei mehrfachen Beobachtungen fur eine Versuchseinheit wird die Reihenfolge der Bedingungen randomisiert. Sowohl für Feld- wie für Laborexperimente (Cook & Campbell, 1979) ist dann eine möglichst eindeutige Interpretation der Variation in den Werten einer oder mehrerer abhangiger Variablen als Folge der Variation einer oder mehrerer unabhängiger Variablen moglich. Bredenkamp (1980) hat in seiner 'Theorie und Planung psychologischer Experimente' eine an der Falsifikationstheorie von Popper orientierte deduktivistische Theorie des Experimentes vorgelegt, die ohne die induktive Verallgemeinerung von Ergebnissen eines Experimentes auskommt (vgl. auch in Ansatzen Dolby, 1982). Hier sollen nur die Aspekte dargestellt werden, die sich direkt mit der Verwendung von Randomisierungs- und Permutationstests unter dem Gesichtspunkt der Sicherung der Validität des statistischen Schlusses für die Prüfung wissenschaftlicher psychologischer Hypothesen befassen. Ansonsten wird auf das Buch von Bredenkamp sowie die Arbeiten

von Bredenkamp (1972, 1980, 1983) und Hager & Westermann (1983 a-c) verwiesen.

Als (deterministische) Allsatze gefaßte psychologische Hypothesen implizieren statistische Populationsaussagen (siehe auch schon Bolles, 1962; Meehl, 1967). Ist nun für die implizierte statistische Hypothese die Fehlerwahrscheinlichkeit für ihr Verwerfen gering, kann als methodische Regel dies als Evidenz für die Falsifizierbarkeit der vorgeordneten wissenschaftlichen Hypothese angesehen werden (vgl. auch Gillies, 1971). Um eine möglichst strenge Prüfung einer wissenschaftlichen Hypothese vornehmen zu können, müssen damit bei Spezifizierung einer minimalen, mit der Wahrscheinlichkeit $1-\beta$ zu entdeckenden Effektstärke α und β klein gewahlt werden. Unter bestimmten Bedingungen (s.u.) laßt sich dann der (minimal) notwendige Stichprobenumfang für das auszuführende Experiment bestimmen. Wichtig ist, daß in diesem Ansatz α und β gleichgewichtig angesehen werden, damit man etwa auch die Möglichkeit hat, eine statistische Nullhypothese als zutreffend ansehen zu konnen, um eine die statistische Alternativhypothese implizierende psychologische Hypothese als nicht bewährt gelten zu lassen. Bei Formulierung statistischer Populationsaussagen werden häufig - wie auch in den zuvor betrachteten linearen Modellen - insbesondere Aussagen über (Lage-)Parameter von uni- oder multivariaten Normalverteilungen in (M)ANOVA-Modellen getroffen. Für die Validitat des gewählten Verfahrens zur statistischen Hypothesenprüfung ist wichtig, daß die experimentellen Beobachtungen an Zufallsstichproben aus der (den) betreffenden Population(en), auf die sich die psychologische

Hypothese bezieht, gewonnen werden. Wie aber verschiedentlich deutlich gemacht wurde (Bredenkamp, 1980; Cox, 1982, Edgington, 1969, 1980; Garfield, 1978; Kirk, 1968) sind in den Humanwissenschaften, insbesondere in der Psychologie interessierende Populationen, sofern sie Gegenstand von als Allsätzen formulierten wissenschaftlichen Hypothesen sind, grundsatzlich offene Populationen. Echte Zufallsstichproben können aber nur aus existierenden, endlichen Populationen gewonnen werden. Wie Fisher (1926, 1935) als erster dargestellt hat (Anscombe, 1948), sind Randomisierungstests für experimentelle Untersuchungen, in denen Versuchseinheiten zufällig verschiedenen experimentellen Bedingungen oder zufälligen Abfolgen einer Reihe von experimentellen Bedingungen zugewiesen werden, valide Tests statistischer (Null-) Hypothesen (vgl. aber die Anmerkungen von Basu (1980) zu späteren Ausgaben der Bücher von Fisher and die nachfolgende Diskussion von Kempthorne (1980) sowie Preece (1983, S. 205f)). Diese Tests sind per Konstruktion verteilungsfreie, bedingte, exakte, finite Tests zum Niveau α. Eine wichtige Annahme ist jedoch die in den zuvor betrachteten linearen Modellen stets vorausgesetzte Additivität der (festen) Behandlungseffekte und der Effekte der experimentellen Beobachtungseinheiten (vgl. Kempthorne, 1980, S. 483; Kempthorne & Barclay, 1953).

Wie in Abschnitt 3 gezeigt wurde, führen das Randomisierungsmodell, in dem die diskreten Verteilungen der Design-Zufallsvariablen (neben den Meßfehlern) das einzige zufallige Element darstellen, und das Permutationstestmodell mit schwachen Vertauschbarkeitsannahmen an die Fehlerzufallsvariablen in den betrachteten Fällen zu identischen Invarianz-Permutationstests. Es ist aber keineswegs so, daß

auch zu jedem parametrischen (M)ANOVA-Testverfahren ein analoger Invarianz-Permutationstest angegeben werden kann, wie das z.B. für den (globalen) Test auf Wechselwirkungen im zweifaktoriellen vollständigen Zufallsplan der Fall ist. In diesem Sinne ist auch eine generelle Empfehlung (Bredenkamp, 1972, 1980, S. 21; Hager & Westermann, 1983a, S. 91), einen F-Test der ANOVA als rechnerische Vereinfachung eines Randomisierungstests zu interpretieren, zu qualifizieren. Kempthorne (1966a, S. 947) zeigt dafür ein Beispiel: In einem Lateinischen Quadrat mit drei Behandlungsstufen und einfacher Zellbesetzung hat ein Randomisierungstest stets das Niveau 1/2, während unter dem üblichen Normalverteilungsmodell der ANOVA jedes Niveau vorkommen kann. Ogawa (1980) gibt andere Beispiele für 4 Stufen.

Ein prinzipielles Problem stellt bei einem bedingten Test, und das ist jeder Permutationstest, die Festlegung des β-Fehlers bzw. der Teststärke vor Ausführung eines Experimentes dar, die aber zur Bestimmung des erforderlichen Gesamtstichprobenumfangs benötigt wird (vgl. Abschnitt 6.1). Man muß dazu die Verteilung der Teststatistik unter der in der (M)ANOVA in einem Effektmaß zusammengefaßten Alternative vor Beginn des Experimentes kennen. Für einen bedingten Test ist aber schon die Verteilung der Teststatistik selbst unter der Nullhypothese trivialerweise nicht vor Ausführung des Experimentes bekannt. Aus diesem Dilemma gibt es mehrere Auswege:

(i) Man setzt auf die in verschiedenen Robustheitsstudien (vgl. Abschnitt 6.4) für einige ein- und zweifaktorielle varianzanalytische Versuchspläne demonstrierte α - und β-Robustheit der F-Statistiken oder anderer zu ihnen in monotoner Beziehung stehenden Teststatistiken unter dem Permutationsmodell gegenüber einzelnen oder kombinierten Verletzungen verschiedener Annahmen des parametrischen Normalverteilungsmodells (Bredenkamp, 1980; S. 21; Hager & Westermann, 1983a, S. 91, 119). Für kleine Stichprobenumfange ist die Annahme der β-Robustheit jedoch auch kritisiert worden (Alf & Abrahams, 1972). Dann konnen (M)ANOVA-Verfahren wiederum, falls ein (Invarianz-)Permutationstest für ein Testproblem existiert, als approximative Testverfahren eingesetzt werden. Eine weitere Stützung erhält dieses Argument dadurch, daß im Normalverteilungsfall die asymptotische relative Effizienz von F-Test und Permutationstest den Wert 1 hat. Bei nicht zu großen Effektstärken und kleinen α-und β-Werten sind die im Normalverteilungsfall erforderlichen Gesamtstichprobenumfänge ebenfalls recht groß, sodaß man hoffen kann, daß auch im finiten Fall der Teststärkeunterschied praktisch vernachlässigbar ist. Allerdings hat Lambert (1985) nachgewiesen, daß der p-Wert des Permutationstests nicht qualitativ robust ist (vgl. Abschnitt 6.4). Folglich sind sicher noch weit umfangreichere Simulationsstudien zur Robustheitsfrage besonders im multivariaten Fall erforderlich.

(ii) Die Identität in der asymptotischen relativen Effizienz im Normalverteilungsfall könnte man auch dazu benutzen, den notwendigen Gesamtstichprobenumfang unter Annahme des parametrischen Normalverteilungsmodells zu berechnen und dennoch einen Permutationstest für die tatsächlich erhobenen Daten auszuführen. Der erforderliche Gesamtstichprobenumfang wird dann in aller Regel so groß sein, daß nur eine Monte-Carlo Lösung (vgl. Abschnitt 8.2) mit einer großen Stichprobe aller unter der jeweiligen Nullhypothese zulässigen Permutationen praktisch trotz effizienter Algorithmen und schneller Rechenanlagen ausführbar ist. Analog zur Empfehlung von Bredenkamp (1980, S. 73f) bei nichtparametrischen Verfahren den notwendigen Stichprobenumfang unter Berücksichtigung der asymptotischen relativen Effizienz der nichtparametrischen Tests im Vergleich mit dem Normalverteilungsanalogon zu berechnen, kann man für Permutationstests folgendermaßen vorgehen. Wie noch in Abschnitt 8.2 näher erläutert wird, nähert sich die Teststärke der Monte-Carlo Lösung 'von unten' der Teststärke für den exakten Permutationstest. Die sog. Dwass-Effizienz (vgl. Abschnitt 8.2, Formel (8.7)) ist ein Maß für die relative Unterlegenheit der Monte-Carlo Lösung in der Teststärke. Als Stichprobenumfang an Beobachtungen für die Monte-Carlo Lösung könnte man dann das durch die Dwass-Effizienz dividierte n aus dem Normalverteilungsfall ansetzen.

(iii) Wenn man berücksichtigt, daß psychologische Hypothesen streng testende sog. Entscheidungsexperimente sinnvollerweise erst nach mehreren Erkundungsexperimenten mit aposteriori-Teststarkebestimmung und Effektschätzung unabhängig von signifikanten oder nicht signifikanten Resul-

taten (Bakan, 1966; Greenwald, 1975; Sahner, 1979; Sterling, 1959; Tullock, 1959) innerhalb eines Forschungsprogramms ausgefuhrt werden, kann man auch an das bereits in Abschnitt 6.3 skizzierte Vorgehen denken. Durch Subtraktion aller (Kleinste-Quadrat-) Effektschätzer in dem jeweiligen linearen Modell von den Beobachtungen eines Erkundungsexperimentes erhält man eine Stichprobe von Residuen, die nur noch Einheiten- und Meßfehler enthält. Bei Replikationen mit derselben (denselben) abhängigen Variablen ergibt sich dann u.U. eine große Stichprobe, die relativ verläßliche Informationen über die Beschaffenheit der Fehlervariablen für die interessierende Population liefert (vgl. aber die kritischen Anmerkungen in Hager & Westermann, 1983a, S. 175f). Wie ebenfalls zuvor erläutert, müssen dann die erwarteten Effekte für alle Stufen der unabhängigen Variable(n) explizit spezifiziert werden. Die Angabe eines bestimmten Anteils der Gesamtvariabilität wie im parametrischen Normalverteilungsmodell ist nicht ausreichend. Durch wiederholtes Ziehen mit Zurücklegen von Zufallsstichproben aus dem 'Residuenpool' von 'realistischem Umfang' (vgl. Taylor & Thompson, 1986) oder in Anlehnung an den Normalverteilungsfall, Addieren der erwarteten Effekte und Ausführen eines entsprechenden Permutationstests kann man in diesem Simulationsexperiment erproben, ob mit dem gewählten Gesamtumfang der Schätzwert für die exakte konditionale Teststärke hinreichend hoch ausfällt. Eventuell muß das Simulationsexperiment mit verändertem Gesamtumfang wiederholt werden, bis ein zufriedenstellendes Ergebnis erzielt wird und das echte Experiment ausgefuhrt werden kann. Der Rechenaufwand für ein solches Vorgehen wird allerdings sehr groß ausfallen (vgl. auch Ab-

schnitt 8.4.3). Man wird somit stets abwägen müssen, ob die Annahme eines parametrischen Modells mit Normalverteilungsannahme für Nullhypothese und (Lage-)Alternative hinreichend gerechtfertigt ist, oder ob nicht ein solch zeit- und kostenaufwendiges Verfahren weniger Storfaktoren (Gadenne, 1976) für eine möglichst strenge Prüfung der psychologischen Hypothese beinhaltet. Andererseits zeigt der vielfach größere Aufwand bei Aufgabe des parametrischen Modells mit Normalverteilungsannahme deutlich, daß man dort 'starke' Modellannahmen getroffen hat (vgl. Kempthorne, 1975, S. 314).

(iv) Ohne den expliziten 'Vorlauf' von 'Erkundungsexperimenten kann man natürlich auch bestimmte Verteilungen mit bekannten Eigenschaften wie größere/kleinere Wölbung als die Normalverteilung oder gewisse Schiefegrade (z.B. bei Reaktionszeitexperimenten) benutzen, um im Simulationsexperiment zu Residuenstichproben zu kommen. Aus diesen lassen sich ebenfalls (Pseudo-) Zufallsstichproben ziehen (vgl. z.B. Kennedy & Gentle, 1980, Kap. 6), und man kann wie unter (iii) vorgehen.

Insgesamt zeigt sich, daß gerade die apriori-Teststärkekontrolle in einem Randomisierungsmodell prinzipielle Probleme aufwirft. Aber auch die Unterstellung eines speziellen Populationsmodells ist mit gravierenden Storungsmöglichkeiten der Validitat des statistischen Schlusses behaftet.

Besonders schwerwiegend erscheinen die oben angeführten Probleme fur multivariate Testprobleme. Im parametrischen Normalverteilungsfall gibt es in der MANOVA bei Verwendung des Spurkriteriums von Pillai, wie Bredenkamp (1980, S. 80ff) und Bredenkamp & Erdfelder (1985) darstellen, keine

zusätzlichen Probleme. Allerdings muß man, um insgesamt nicht zu sehr an Teststarke einzubußen, davon ausgehen, daß die Einflusse der unabhängigen Variable(n) sich, wenn auch nicht unbedingt in gleichem Maße, auf alle betrachteten abhängigen Variablen auswirken. Simulationsstudien der Effekte, wie in (iii) oder (iv) mit Spezifizierung der Effekte für jede Stufe der unabhängigen Variable(n) und jede abhangige Variable erscheinen kaum realistisch durchführbar, und auch der Rechenaufwand würde, da jeder einzelne Permutationstest für eine Zufallsstichprobe viel mehr Rechenzeit als im univariaten Fall benotigt (vgl. Abschnitt 8.4), im allgemeinen unvertretbar hoch ausfallen. Damit bleiben nur die Auswege (i) oder (ii) übrig.

7.2 Bedeutung für die Legitimierung statistischer Hypothesentests

Auch wenn man der deduktivistischen Theorie des Experimentierens nicht folgt, sondern induktivistisch argumentiert, sind Randomisierungstests ein wichtiges Hilfsmittel, wie auch Bredenkamp (1980, S. 18f) ausführt. Ebenso verhält es sich für Erkundungsexperimente (vgl. O'Neill & Wetherill, 1971, sog. fundamental exploration) zu Beginn eines Forschungsprogrammes, wenn man noch nicht fundiert bestimmte Verteilungsannahmen machen kann. In beiden Fällen sind echte Zufallsstichproben in der experimentellen psychologischen Forschung kaum aufzufinden (vgl. auch Molenaar, 1985, S. 178f; Kempthorne, 1977, S. 5f). In der statistischen Literatur ist dieser Aspekt der validen Testung statistischer

Hypothesen (ausschließlich) aufgrund experimenteller Randomisierung häufiger diskutiert worden (Fisher, 1955; Rubin, 1978, Wu, 1981; Neyman , 1967, S. 1459; Yates, 1964, 1975). Der induktive Schluß von den untersuchten auf andere Beobachtungseinheiten derselben Population wurde nicht mehr als statistische, sondern 'wissenschaftliche' (scientific) Inferenz angesehen (vgl. z.B. Edgington, 1980, S. 5; Kempthorne, 1952, S. 152, 1971; Welch, 1937, S. 47). Zu beachten ist jedoch die Unmöglichkeit der Bestätigung einer empirischen Hypothese (z.B. Mahoney, 1978; Meehl, 1967, 1978; Lykken, 1968). Allein für die in einem Experiment berucksichtigten Beobachtungseinheiten, die zu einer bestimmten Population gehören mögen, wird dann geprüft, ob verschiedene experimentelle Bedingungen die abhängige(n) Variable(n) in gleichem Ausmaß beeinflussen. Entsprechend wird das Resultat der Permutationstests nicht auf eine Population verallgemeinert.

Als ein zweiter wichtiger Aspekt dieses Ansatzes wird die enge Verbindung zwischen tatsächlichem experimentellem Versuchsplan und verwendetem statistischen Modell angesehen, während in der üblichen (M)ANOVA ein Modell 'einfach angenommen' (Zyskind, 1975) wird.

Auch wenn R.A. Fisher dem Teststärkekonzept ablehnend gegenüberstand (1935, S. 17), ist doch in der Folgezeit dem Aspekt der Sensitivität und Optimalität (Dempster & Schatzoff, 1965; Robinson, 1979) von Permutationstests durchaus Bedeutung beigemessen worden.

Obwohl an verschiedenen Stellen dieser Arbeit schon auf diese drei Gesichtspunkte eingegangen worden ist, sollen hier noch einmal einige zusammenfassende Bemerkungen angeführt werden.

(i) Gültigkeit des Signifikanztests aufgrund experimenteller Randomisierung:

Edgington (1980, Kap. 1.9) nennt R.A. Fisher (1926, 1935) und O. Kempthorne (1952) als diejenigen, die die Bedeutung der experimentellen Randomisierung am deutlichsten hervorgehoben haben (Anscombe, 1948, S. 188f; Kempthorne, 1966, 1972, S. 185, 1975, 1977; Kruskal, 1980). Nicht nur die Annahme einer bestimmten Verteilung ist nicht erforderlich, sondern auch nicht die Annahme von Zufallsstichproben aus spezifizierten Populationen (vgl. Cochran & Cox, 1957, S. 7; Gabriel 1979; Gabriel & Hall, 1983; Greenberg, 1957; Gridgemann, 1959; Lindley & Novick, 1981; Scheffe, 1959; Kempthorne, 1972; Rao, 1965, S. 423). Allein die probabilistischen Eigenschaften der Versuchsplan-Zufallsvariablen, die die randomisierte Zuteilung von Einheiten zu Stufen der unabhängigen Variable(n) widerspiegeln, reichen für die Konstruktion eines statistischen Tests aus (Bunke & Bunke, 1978; Edgington, 1975a; Scheffe, 1959; Kempthorne, 1966). Wilk (1955) bezeichnet die statistische Kontrolle der Einheitenfehler als Hauptfunktion der Randomisierung. Sie ist entscheidend für die interne Validität eines Experimentes (Keppel, 1973, S. 313). Wie in Abschnitt 3.3 ausführlicher dargestellt, gibt es in diesen (additiven) sog. 'Fisher-Modellen' keine Meßfehler. Die Einbeziehung solcher i.a. zufälliger Variationen in den Ausprägungen der abhängigen Variable(n) ist für die Psychologie aber sicher notwendig. Schon Neyman (1935), McCarthy (1939) und besonders Ogawa (1961, 1963, 1974, 1980), Okuno (1958, 1962) und Wehrhahn (1986) sowie Wehrhahn & Ogawa (1982) haben fur verschiedene Testprobleme und Versuchsplane nachgewiesen, daß die Hinzunahme unabhängiger (normalverteilter) Meßfehler

keine zusätzlichen Probleme ergibt. Dasselbe gilt für die in dieser Arbeit dargestellten Invarianz-Permutationstests. Die Annahme einer Verteilung für die Meßfehler führt zu keinem 'Populationsproblem' für die Beobachtungseinheiten. Die Voraussetzung einer Population von moglichen Ausprägungen der Meßfehlervariablen kann man ohne Schwierigkeiten ausschließlich für jede einzelne der am Experiment beteiligten Beobachtungseinheiten treffen. Nur wenn keine Einheitenfehler auftreten können, hat die Sichtweise von Harville (1975, vgl. auch Basu, 1980 und Thornett, 1982a, b), daß experimentelle Randomisierung überflüssig sei, eine gewisse Berechtigung (vgl. auch Lorenzen, 1984; Wilk, 1955, S. 70).

(ii) Verbindung zwischen experimenteller Struktur und mathematischem Modell:

Über die Entscheidung zwischen angenommenen gegenüber abgeleiteten Modellen hinaus (vgl. Wilk, 1955 und Wilk & Kempthorne, 1956, 1957) sind von verschiedenen Autoren genauere Darstellungen der Beziehung zwischen experimenteller Randomisierung und verwendetem Versuchsplan vorgelegt worden (Brien, 1983; Nelder, 1964a, b; Finch, 1979, 1982; Thornett, 1982a,b; White, 1975). Ohne die unterschiedlichen Begriffsbildungen zu referieren, sollen lediglich einige Gesichtspunkte erläutert werden. Ein Versuchsplan, wie er z.B. in dem Buch von Kirk (1968) für alle gängigen Plane schematisch dargestellt ist, gibt keine erschöpfende Auskunft über tatsächlich in einem Experiment ausgeführte zufällige Zuweisungen von experimentellen Einheiten ('experimental units' bei White, 1975) zu den Platzen oder 'Stellen' des experimentellen Planes ('experimental design units'). Wenn z.B. in einem Split-Plot Plan der Plotfaktor nicht für die Stufen einer

vom Versuchsleiter manipulierbaren unabhangigen Variablen steht, sondern verschiedene Auspragungen einer organismischen (naturlichen) Variablen wie Geschlecht, Altersgruppe, etc. darstellt, hingegen der Splitfaktor von verschiedenen in der Abfolge randomisierten, experimentellen Bedingungen gebildet wird, kann die Validitat eines Permutationstests auf Wechselwirkungen zwischen den beiden unabhängigen Variablen nicht auf die experimentelle Randomisierung gegrundet werden. Dann muß man annehmen, daß die Beobachtungseinheiten zufallig aus einer (hypothetischen) Population gewonnen sind. Man kann dann auch einen Permutationstest ausfuhren, weil aus der gerade erwahnten Voraussetzung bei der Gultigkeit der Wechselwirkungsnullhypothese die Permutationsinvarianz der Einheiten uber alle Plotstufen folgt (vgl. auch Robinson, 1980). Die Validität des statistischen Schlusses ist aber einer großeren Gefahrdung ausgesetzt. Ahnlich verhält es sich, wenn ein Faktor mit Meßwiederholungen verschiedene Beobachtungszeitpunkte reprasentiert. Die unterschiedlichen Permutationsstrukturen für den Split-Faktor zeigen sich z.B. auch in den unterschiedlichen Kovarianzen zwischen zwei Beobachtungen im Split-Plot Plan bzw. Split-Split-Plot Plan (vgl. Abschnitt 3.1.5 (i) bzw. (xiii)) je nachdem, ob experimentell randomisiert wurde oder nicht.

(iii) Das Teststärkekonzept und Randomisierungs/ Permutationstests:

R.A. Fisher verwendete Randomisierungstests wie auch andere statistische Signifikanztests, um Evidenz gegen die jeweilige Nullhypothese zu gewinnen (vgl. z.B. Kempthorne, 1971, S. 480f). Seine Auseinandersetzung mit der Neyman-Pearson Testtheorie des Hypothesentestens hat viele Diskus-

sionen initiiert (vgl. z.B. Cox, 1958; Fisher, 1955; Lehmann, 1959; Pearson, 1937, S. 64; Savage, 1976; Tukey, 1960), sodaß z.B. noch 1976 ein ganzes Heft der Communications in Statistics - Theory and Methods dem Vergleich von 'Signifikanztests' und 'Hypopthesentests' mit Beiträgen von Kempthorne, Neyman, Pratt und Roberts gewidmet ist (Kadane, 1976). Auch in der psychologischen Methodenliteratur findet man viele Bezuge auf dieses Thema (z.B. Bredenkamp. 1972; Harnatt, 1975; Morrison & Henkel, 1970). An dieser Stelle sollen allerdings nur noch einmal für Randomisierungs-und Permutationstests relevante Aspekte angesprochen werden.

Wie im parametrischen Fall ist der p-Wert allgemein kein sinnvoll über verschiedene Datensatze vergleichbares Maß für die Stärke eines Effektes bzw. das Ausmaß der Abweichung von der jeweiligen Nullhypothese (Gibbons & Pratt, 1975; Pratt & Gibbons, 1981; de Soete, 1982). Der p-Wert ist Realisation einer Statistik, die unter der Nullhypothese eine Gleichverteilung und eine andere Verteilung unter einer Alternative besitzt (Kempthorne, 1976, 1977). Robinson (1979) hat, wie bereits erwähnt, Definitionen für einen optimalen Test gegeben, die den Bezug auf das Akzeptieren oder Verwerfen von Hypothesen nicht benötigen und statt Teststärke von 'Sensitivität eines Tests' sprechen. Wie in Abschnitt 6.3 demonstriert, bereitet es keine prinzipielle Schwierigkeit, für die im Experiment erhaltenen Daten die exakte bedingte Teststärke (Sensitivität) bei Wahl einer bestimmten Teststatistik gegen jede (Lage-)Alternative zu bestimmen (vgl. Kempthorne, 1977, S. 14f). In Ein-und Zweistichprobenproblemen kann sogar die gesamte Teststärkekurve gegen Verschie-

bealternativen ermittelt werden. Mathematisch nicht handhabbar (vgl. Kempthorne, 1977, S. 18f) scheint aber das Problem zu sein, welche Teststatistik zu einem besonders sensitiven Permutationstest führt. Wie an verschiedenen Stellen der Arbeit erwähnt, sind die Teststatistiken aus der Theorie des linearen Modells mit Normalverteilungsannahme 'gute Kandidaten', und in der dargestellten Theorie der Invarianz-Permutationstests werden sie im wesentlichen auch verwendet. Besonders für einfaktorielle Fragestellungen ist aber eine Fülle von anderen Teststatistiken denkbar und auch bereits vorgeschlagen worden (vgl. Birnbaum, 1970; Pratt & Gibbons, 1981, Kap. 4.6). Insbesondere bei Zweifel an der Intervallskaliertheit einer abhängigen Variablen ist die Differenz der (gegen monotone Transformationen invarianten) Mediane (vgl. schon Nair, 1940; Kempthorne, 1977; Oetting, 1965) im unverbundenen Zweistichprobenproblem eine geeignete Teststatistik. Für multivariate abhängige Variablen sind,wie bereits angeführt, die Zweifel an einer Übertragbarkeit der Teststatistiken aus dem Normalverteilungsfall größer (Kempthorne, 1966b, 1977) und für die MRPP-Statistiken (vgl. Abschnitt 4.7) wurde bereits nachgewiesen, daß nichtnormale, schiefe asymptotische Verteilungen der verwendeten Teststatistiken resultieren können.

Für (Wechselwirkungs-)Testprobleme in mehrfaktoriellen linearen Modellen erscheinen die Möglichkeiten einer sinnvollen Teststatistikauswahl jedoch eingeschränkter zu sein. Wie es auch bei Rangtests in solchen Modellen (vgl. Erdfelder & Bredenkamp, 1984 für einen Überblick) schwerwiegende Probleme gibt, Test zu konstruieren, die unabhängig von den nicht interessierenden (störenden) Parametern sind, sind auch die Vorschläge von Still & White (1981) für belie-

bige lineare Modelle zweifelhaft. Wie zuvor gezeigt, ist nicht für alle Testprobleme deren Invarianz zu sichern, wenn störende Parameter bzw. deren Schätzungen wegtransformiert werden und der Permutationstest für die transformierten Beobachtungen ausgeführt wird. Einen ähnlichen Fehler begehen Collier, Baker & Mandeville (1967), die für einen Test auf Wechselwirkungen im Split-Plot Plan alle unter der globalen Nullhypothese (keine Haupt-und Wechselwirkungseffekte) zulässigen Permutationen betrachten und eine 'Überlegenheit' des Permutationstests gegenüber dem F-Test in bezug auf die Teststärke feststellen. Wie zuvor erwähnt, ist es leicht, für das univariate Ein-und Zweistichprobenproblem den bedingt und unbedingt teststärksten Permutationstest gegen jede vollständig spezifizierte Alternative und nicht nur eine Lageverschiebung im Normalverteilungsfall anzugeben (vgl. schon Lehmann & Stein, 1949; Box & Andersen, 1955 sowie Pratt & Gibbons, 1981, Kap. 4.4 und 6.4). Dazu setzt man die beobachteten Werte und alle zulässigen Permutationen in die alternative Dichtefunktion oder diskrete Wahrscheinlichkeitsverteilung ein. Der $[\alpha|Q_n|]$ -te Anteil bestehend aus den größten Werten dieser Dichte bzw. Wahrscheinlichkeitsverteilung in absteigender Reihenfolge geordnet bildet dann die kritische Region des Tests. Der p-Wert für den Test der Nullhypothese: 'Keine Unterschiede'gibt die Anzahl an zulassigen Permutationen mit nicht weniger extremen (großeren) Funktionswerten als für die tatsächlichen Beobachtungen an. Für die Planung des minimal notwendigen Gesamtstichprobenumfanges in einem Experiment mit fixiertem α und β -Fehler sowie explizit oder über den Anteil erklärter Varianz fixierter (Lage-)Alternative sind die bisherigen Ausführun-

gen wie die in Abschnitt 6 prinzipiell nicht ausreichend. Ohne eine explizite Spezifizierung einer multivariaten Verteilung über den Lageparameter hinaus ist die Teststärke apriori nicht exakt bestimmbar. Wie in Abschnitt 7.1 vorgeschlagen, bleibt es weiterhin sicherlich aufwendigen Simulationen vorbehalten zu zeigen, inwieweit die approximativen Stichprobenumfangsbestimmung anhand des Normalverteilungsfalles oder anhand von diskreten Verteilungen größerer Residuenstichproben erfolgreich ist. Dies ist insbesondere für multivariate abhängige Variable der Fall, auch wenn man die grundsätzlichen Probleme von Kempthorne (1966b) mit multivariaten Daten in experimentellen Untersuchungen berücksichtigt.

7.3 Das Testen der Effekte aller unabhängigen Variablen in uni- und multivariaten Versuchsplänen

7.3.1 Problematische Aspekte dieses Vorgehens

In vielen psychologischen Lehrbüchern zur Versuchsplanung und -auswertung wird wie in der entsprechenden statistischen Methodenliteratur wenig berücksichtigt, daß es in einem mehrfaktoriellen Versuchsplan bei Ausführung aller (varianzanalytischen) Tests auf (einfache) Haupt- und (einfache) Wechselwirkungen zur Erhöhung der gesamten α- und β -Fehler im Vergleich zum nominellen α-und β-Niveau der einzelnen Signifikanztests kommt. Zur Vergabe des Fehlerniveaus sind unterschiedlichste Vorschläge gemacht worden. Es wurden verschiedene Testfamilien definiert, für die global eine Irrtumswahrscheinlichkeit festzulegen ist. Sie reichen von

der Zusammenfassung aller Tests eines Experimentes über die Zusammenfassung aller Tests bezüglich eines Faktors bis hin zur separaten Betrachtung eines jeden Einzeltests (siehe z.B. für einen Überblick Hager & Westermann, 1983a, Kap. 8.3 und 8.4; Westermann, 1982). Beispiele für solche Testfamilien, insbesondere sog. multiple Tests, wurden in Abschnitt 5 behandelt.

Bei der Analyse von F-Statistiken in Varianzanalysetafeln ist deren Abhängigkeit zu berücksichtigen, sofern verschiedene F-Brüche dieselbe Fehlerquadratsumme als Nenner enthalten (Kimball, 1951; Hays, 1963; Hurlburt & Spiegel, 1976; Sonnemann, 1982). Dieselbe Beziehung für mindestens ein fälschliches Verwerfen einer Nullhypothese bei k Tests mit nominellen Niveaus α_i

(7.1)
$$\alpha < 1 - \prod_{i=1}^{k} (1-\alpha_i)$$

gilt auch für die analogen MANOVA-Teststatistiken, selbst wenn für die Einzeltests verschiedene der Teststatistiken aus (4.18) verwendet werden, sofern nur der Fehlerterm identisch ist (Dykstra, 1979).

Allen diesen verschiedenen Teststrategien ist darüberhinaus gemeinsam, daß oft keine direkte Entsprechung zwischen der inhaltlichen Fragestellung und den mit den statistischen Verfahren getesteten statistischen Null- und besonders Alternativhypothesen besteht. Das ist insbesondere dann problematisch, wenn die statistischen Testverfahren nicht zu explorativen Zwecken sondern zur Entscheidung über wissenschaftliche Hypothesen mit herangezogen werden. Wie bereits erwähnt, haben besonders Bredenkamp (1969, 1972, 1980) und Hager & Westermann (1983a) sowie Westermann & Hager (1983

a,b) Vorgehensweisen vorgeschlagen, die einen moglichst direkten und engen Zusammenhang zwischen interessierender(en) wissenschaftlicher(en) Hypothese(n) und tatsächlicher getesteter(en) statistischer(en) Hypopthese(n) herstellen. Weitere Aspekte der Einsatzmöglichkeiten von Permutationstests in diesem Ansatz werden in Abschnitt 7.4 dargestellt. Zuvor soll eine 'konventionelle' Analyse wie in Kirk (1968) für uni- oder multivariate Beobachtungen in einem mehrfaktoriellen Versuchsplan anhand von Invarianz-Permutationstests dargestellt werden.

7.3.2 Univariate mehrfaktorielle Experimente

Das Vorgehen wird vorwiegend am Beispiel des Split-Plot Versuchsplans erläutert. Willmes & Pyhel (1981) und Willmes (1982) haben eine Prozedur vorgeschlagen, die aus den Schritten (i) Wechselwirkungseffekte, (ii) (einfache) Haupteffekte und (iii) multiple Vergleiche von Paaren von Faktorstufen besteht und je Schritt in (ii) und (iii) zusätzlich je unabhängiger Variabler jeweils das (multiple) Fehlerniveau 1. Art bei α kontrolliert. Der jeweilige β -Fehler wurde nicht explizit festgelegt.

Da je nach zu testender Hypothese und je nach vorab durch Invarianzbetrachtungen zu eliminierenden störenden Effekten u.U. unterschiedliche Permutationsinvarianzstrukturen der jeweiligen Teststatistiken unter der jeweiligen Nullhypothese vorliegen, kann nicht (ohne großen überflüssigen Rechenaufwand für Datenpermutationen, s.u.) wie in der parametrischen ANOVA in einem Rechengang durch Bestimmung aller F- oder analoger Teststatistikwerte über die Hypothesen ent-

schieden werden. Für Tests auf Wechselwirkungen (und Plot-Haupteffekte) sind Vertauschungen der Blöcke zulässig, für die Analyse eines Split-Effektes mußten aber Permutationen innerhalb der Blöcke und nicht zwischen (ganzen) Blöcken durchgeführt werden. Sinnvolle Teststatistiken für je einen der Haupteffekte sind aber invariant unter zulässigen Permutationen für den Test auf den anderen Haupteffekt. Falls keine Wechselwirkungen vorliegen oder man zumindest hinreichend sicher von dieser Annahme ausgehen kann, müssen zum Testen des Plotfaktors mit der Notation aus Abschnitt 3.1.5 (xi) die Summen bzw. Mittelwerte über die Split-Stufen je Block betrachtet werden. Diese $\bar{Y}_{ih(i).}$ werden wie die Zufallsvariablen im unverbundenen k-Stichprobenproblem in 3.1.5 (iii) behandelt und

$$(7.2) \qquad S_A(Y) = \sum_{i=1}^{r} n_i \bar{Y}^2_{i..}$$

ist eine geeignete Teststatistik. Diese ist unter den $n.!/\prod_{i=1}^{r} n_i!$ (zulässigen) Permutationen der $\bar{Y}_{ih(i).}$ zu betrachten.

Für einen Test des Split-Faktors ist generell nur das Modell mit der Invarianzstruktur $S_c \sim S_{n.}$ geeignet, d.h. die Fehler-Zufallsvariablen je Block müssen vertauschbar sein. Ohne Wechselwirkung können dann alle n. Blöcke als Replikationen aufgefaßt werden und man wählt die Teststatistik des verbundenen k-Stichprobenproblems aus 3.1.5 (ii) für die um die Blockmittel korrigierten Beobachtungen $Y_{ih(i)j} - \bar{Y}_{ih(i).}$. Da diese Blockmittel S_c-invariant sind, kann man die Teststatistik in folgender Form unter den $(c!)^{n.}$ zulässigen Intrablockpermutationen betrachten:

$$(7.3) \qquad S_B(Y) = \sum_{j=1}^{c} \bar{Y}^2_{..j} \quad .$$

Zur Analyse weiterer Unterschiede zwischen Faktorstufen von A bzw. B kann die in Abschnitt 5.4.1 beschriebene sequentielle multiple Testprozedur nach Peritz verwendet werden. Diese ist teststärker als das in den oben zitierten Arbeiten verwendete Verfahren von Holm (1979) oder dessen Verschärfungen (Hommel, 1985; Shaffer, 1984, 1986).

Im Fall von Wechselwirkungen kann man wie bei Kirk (1968, Kap. 8.6) vorgehen. Es werden c bzw. r Permutationstests für das unverbundene bzw. verbundene k-Stichprobenproblem ausgeführt und das nominelle Niveau für den Fehler 1. Art je Test auf α/c bzw. α/r festgesetzt (vgl. Willmes & Pyhel, 1981). Auf die Unterscheidung verschiedener Typen von Interaktionen (vgl. z.B. Bredenkamp, 1980, S. 24ff) und unterschiedliche Konsequenzen für die Wahl eines Tests auf (einfache) Haupteffekte wird in der statistischen Versuchsplanungsliteratur i.a. wie auch bei Kirk (1968) nicht näher eingegangen (vgl. Abschnitt 7.4). Bei signifikantem einfachem Haupteffekt kann dann jeweils das Peritz-Verfahren zum multiplen Niveau α/c bzw. α/r durchgeführt werden.

Falls nur bezüglich des Plot-Faktors eine experimentelle Randomisierung möglich ist, kann man neben dem univariaten Test auf Profilparallelität mit Teststatistik (3.65) bzw. (3.67) oder (3.68) auch eine der mit (4.57) gebildeten multivariaten Permutationsteststatistiken wählen. Neben diesem globalen Test kann für zweiseitige Fragestellungen u.U. die in Abschnitt 5.3 vorgestellte multiple Testprozedur für die (c-1) Dimensionen der Kontrastvektoren Z_{ih} aus (4.57) Verwendung finden, um eine Verletzung der Profilparallelität genauer zu 'lokalisieren'. Welche Stufen des Plot-Faktors zu der fehlenden Parallelität beitragen, läßt sich mit dem Peritz-Verfahren für die (c-1)-dimensionalen

Kontrastvektoren feststellen.

Falls nur die Stufen des Split-Faktors den Beobachtungseinheiten zufällig zugewiesen werden können, die Einheiten selbst aber lediglich bezüglich einer organismischen Variablen geschichtet (stratifiziert) sind, ist im strengen Sinne für die Wechselwirkung kein Permutationstest in einem Randomisierungsmodell möglich. Wie bereits erläutert, muß man ansonsten annehmen, daß die Einheiten wie Zufallsstichproben aus hypothetischen Populationen angesehen werden können, für die die Fehler im 'verteilungsfreien' linearen Modell die bekannten Invarianzeigenschaften von zuvor haben. Ansonsten ist wie für den zweifaktoriellen vollständigen Zufallsplan nur die Hypothese testbar, daß der Splitfaktor 'überhaupt keinen', d.h. auf keiner Stufe des Plotfaktors einen Effekt hat. Als Teststatistik ergibt sich dann nach Eliminierung der störenden Blockeffekte durch

$$T_{ihj} := (KY)_{ihj} - \bar{Y}_{ih.} \quad \text{mit} \quad K = I_n \otimes (I_c - 1/c J_c) \quad :$$

$$(7.4) \qquad S_{B,AB}(T) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_i \bar{T}^2_{i.j} \quad .$$

Da $\bar{Y}_{i..}$ invariant ist unter den bei Gültigkeit der Nullhypothese zulässigen Intrablock-Permutationen aus $\Omega_n = \times_{i=1}^{r} \times_{h=1}^{n_i} S_c$ könnte aquivalent auch

$$(7.5) \qquad S_{B,AB}(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} n_i \bar{Y}^2_{i.j}$$

benutzt werden. K ist auch gegen die großere Gruppe $\Omega^*_N = S_c \wr S_{n.}$ invariant. Da aber bezuglich des Plotfaktors keine experimentelle Randomisierung vorliegen sollte, ist als Permutationsgruppe für die Fehlerverteilung eher ebenfalls Ω_N als Ω_N^* oder auch $\Omega_N^{**} = \times_{i=1}^{r} (S_c \wr S_{n_i})$ plausibel.

Abschließend soll noch an mehreren Beispielen erläutert werden, wie in Plänen mit wenigstens zwei experimentellen unabhängigen Variablen getestet werden kann, ob überhaupt

ein Effekt dieser Faktoren und ihrer Wechselwirkungen vorliegt (vgl. Bredenkamp, 1980, S. 64f, 67f, 93ff).

Im Gegensatz zum Test auf Wechselwirkung im CRF-rc ist dieser globale Test mit einem Invarianz-Permutationstest möglich. Als Modellgleichung hat man in der Zellmittelwertsnotation wie z.B. bei Krafft (1978, S. 79ff):

$$Y_{ijk} = a_{ij} + \varepsilon_{ijk} \quad ,i=1,..,r;\ j=1,..,c;\ k=1,..,n_{ij},$$

$$N=\sum_{ij}\sum n_{ij} .$$

In einem Experiment mit zufälliger Zuteilung zu den Stufen der beiden Faktoren A und B sei o.B.d.A. $n_{ij}=n>1$ und der Vektor ε verteilt nach S_N-invariantem $F(x), x\in R^N$.

Wenn keine Effekte der beiden unabhängigen Variablen vorliegen, sind die Beobachtungen aus allen rc Zellen des Versuchsplanes beliebig miteinander vertauschbar. Als Teststatistik hat man

$$(7.6) \qquad S_{A,B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c} \bar{Y}^2_{ij.} \quad ,$$

deren Verteilung unter allen $N!/(n!)^{rc}$ Permutationen zu betrachten ist. Bei ungleichen Zellbesetzungen, steht in (7.6) noch der multiplikative Faktor n_{ij}.

Dieser Globaltest aus (7.6) ist in offensichtlicher Weise auf Kreuzklassifikationen mit mehr als zwei Faktoren erweiterbar.

Im zweifaktoriellen Blockplan mit (vgl. auch (3.88)):

$$Y_{hij} = a_{ij} + \pi_h + \varepsilon_{hij} \quad ,i=1,..,r;\ j=1,..,c;\ h=1,..,n,$$

und (mindestens) $Q_N=\times_{h=1}^{n} S_{rc}$ -invariantem Fehlervektor (vgl. 4.69) eliminiert man die für den Test H: $a_{ij}=0$ für alle i,j störenden Blockeffekte durch $K=I_n\otimes(I_{rc}-1/rcJ_{rc})$:

$$T_{hij} = (KY)_{hij} = Y_{hij} - \bar{Y}_{h..} \quad .$$

Als Teststatistik eignet sich dann

$$(7.7) \qquad S_{A,B,AB}(T) = \sum_{i=1}^{r}\sum_{j=1}^{c} n\bar{T}^2_{.ij}$$

oder äquivalent, da $\bar{Y}_{h..}$ $S_{rc} \sim S_n$-invariant ist:

(7.8) $$S_{A,B,AB}(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{Y}^2_{.ij} \quad .$$

Für den exakten Invarianz-Permutationstests ist diese Teststatistik unter allen $((rc)!)^n$ Intrablockpermutationen zu berechnen.

Auch dieser Globaltest mit der Teststatistik aus (7.8) ist leicht auf mehr als zwei intraindividuelle, gekreuzte experimentell variierte unabhängige Variablen verallgemeinerbar. Es sei noch angemerkt, daß bei fehlenden Wechselwirkungen Tests auf Haupteffekte jedes der beiden Faktoren mit der Teststatistik für das verbundene k-Stichprobenproblem auszuführen sind, wenn dort die Intrablocksummen (oder Mittelwerte) $Y_{hi.}$ bzw. $Y_{h.j}$ über die Stufen des jweils zweiten Faktors als Zufallsvariablen der Beobachtungen angesehen werden. Analoges gilt für den CRF-rc.

In dreifaktoriellen Plänen, die Faktoren ohne und mit Meßwiederholungen enthalten, kann man wie im Split-Plot Plan unterscheiden zwischen intra- und interindividuellen Faktoren. Wie bei Bredenkamp (1980, S. 93ff) kann man nun global testen, ob irgendwelche inter - bzw. intraindividuellen Effekte vorhanden sind. Ein globaler Test auf Abwesenheit aller experimentellen Effekte ist schon im Split-Plot Plan mit festen Blockeffekten nicht möglich. Es gibt keine Transformationsmatrix K für $Y_{ih(i)j}$ aus (3.61), als deren Ergebnis alle Effekte außer dem Blockeffekt π_h übrigbleiben.

Im Split-Split-Plot Plan überlegt man für den Test auf Fehlen aller intraindividuellen Effekte Folgendes:

In der Modellgleichung (3.79) sind bis auf α_i alle Modellparameter intraindividuelle Parameter. Wie für den Split-Plot Plan bleiben nach Subtraktion der Blockmittel nur noch Parameter für intraindividuell variierte experimentelle Faktoren übrig. Die zugehörige Transformationsmatrix K mit $(KY)_{ihjk}=Y_{ihjk}-\bar{Y}_{ih..}$ ist das $S_{cs}\sim S_n$.-invariante $K=I_n \otimes (I_{cs}-1/csJ_{cs})$. Als Teststatistik hat man dann :

(7.9) $$S_{B,C,BC,AB,AC,ABC}(T) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n_i \bar{T}^2_{i.jk} \quad .$$

Diese wird unter allen Intrablockpermutationen betrachtet. Da $\bar{Y}_{i...}$ unter diesen Permutationen invariant ist, reicht es auch,

(7.10) $$S_{B,C,BC,AB,AC,ABC}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{s} n_i \bar{Y}^2_{i.jk}$$

unter allen $((cs)!)^{n.}$ Intrablockpermutationen zu berechnen. Vorschläge zur Effektschätzung bei Faktoren mit wiederholten Messungen machen z.B. Gaebelein & Soderquist (1978); Bredenkamp (1980).

Zum Testen des Plotfaktors, falls keine zweifachen Wechselwirkungen mit Beteiligung des Plotfaktors und keine dreifache Wechselwirkung vorliegen, wird für die Intrablockmittel $\bar{Y}_{ih..}$ wie im CRF-k Verfahren

(7.11) $$S_A(Y) = \sum_{i=1}^{r} n_i \bar{Y}^2_{i...}$$

als Teststatistik gewählt.

Eine Analyse aller Effekte im Split-Split-Plot Plan in Analogie zum Split-Plot Plan ist ebenfalls im Sinne von Kirk (1968, Kap. 8.14) ohne Probleme möglich. Bei Vorliegen einer dreifachen Wechselwirkung sind für die Tests auf einfache Interaktionen r-mal wie im RBF-cs für jede Stufe des Plot-

faktors und jeweils s- bzw. c-mal wie im Split-Plot Plan für jede Stufe des anderen Splitfaktors C bzw. B Invarianz-Permutationstests auf Wechselwirkung auszuführen. Die Analyse einfacher, einfacher Haupteffekte ist dann bei signifikantem Ausgang dieser Tests auf reduziertem Niveau mit Tests für (un-)verbundene k-Stichprobenprobleme lösbar. Ohne dreifache Wechselwirkung wird für die Wechselwirkung der Splitfaktoren wie im RBF-cs Plan mit nun insgesamt n. Blöcken verfahren. Für die Tests auf zweifache Wechselwirkungen zwischen einem Split- und einem Plotfaktor wird wie im Split-Plot Plan für die Zufallsvariablen $\bar{Y}_{ihj.}$ bzw. $\bar{Y}_{ih.k}$ analysiert. Die anschließenden Tests auf (einfache) Haupteffekte sind dann ebenfalls offensichtlich. Auf die Tests der verallgemeinerten Profilparallelitat wie in Abschnitt (4.6) und die Möglichkeiten einer multiplen Testprozedur für zweiseitige Tests für die Komponenten der (c-1)(s-1)-dimensionalen Kontrastvektoren in Analogie zum Profilparallelitätsproblem im Split-Plot Plan sei nur hingewiesen.

Im Split-Plot-Plot Plan ist ein globaler Test auf fehlende Effekte der beiden Plotfaktoren nur möglich, wenn man eine verschwindende dreifache Interaktion und verschwindende zweifache Wechselwirkung zwischen dem Splitfaktor und jedem der beiden Plotfaktoren einzeln annehmen kann. In diesem Fall ist wie zuvor für den Globaltest im CRF-rc jetzt mit den Zufallsvariablen $\bar{Y}_{ijh.}$ zu verfahren. Eine vollständige Analyse dieses Versuchsplanes ist ebenfalls nur mit Einschränkungen moglich. Invarianz-Permutationstests auf eine (einfache) Wechselwirkung der beiden Plotfaktoren müssen wegen der Nichtexistenz eines solchen Tests fur den CRF-rc

ebenfalls unmöglich sein. Die (einfachen) Wechselwirkungen zwischen dem Split- und je einem Plotfaktor können dagegen getestet werden.

Für alle in diesem Abschnitt vorgeschlagenen Invarianz-Permutationstests sind neben der Kontrolle des α-Niveaus die Überlegungen zur approximativen apriori-Festlegung des benötigten Gesamtstichprobenumfangs unter Spezifizierung des β-Fehlers und eines (minimalen) erwarteten Effektes aus Abschnitt 7.1 (i) und (ii) ebenfalls gültig. Die Vorgehensweise im Normalverteilungsfall ist in Bredenkamp (1980) ausführlich dargestellt.

7.3.3 Multivariate mehrfaktorielle Experimente

Wie bei der Verallgemeinerung der (Invarianz-)Permutationstests aus Abschnitt 3 auf multivariate Beobachtungen in Abschnitt 4 sind auch alle Verfahrensvorschläge aus Abschnitt 7.3.2 übertragbar. Die Summen von quadrierten Größen in den Formeln (7.2) - (7.11) sind wieder durch dyadische Produkte d-dimensionaler Vektoren zu ersetzen. Wie zuvor sind z.T. in den Formeln noch permutationsinvariante Größen hinzuzufügen, um die dxd-dimensionalen $\underline{S}_h$ und $\underline{S}_e$-Matrizen zu bilden. In (7.2) hat man zum Testen des Plotfaktors im d-dimensionalen Split-Plot Plan wie im d-dimensionalen k-Stichprobenproblem in (4.29):

$$(7.12)\quad \underline{S}_h(Y) = \sum_{i=1}^{r} n_i (\bar{\underline{Y}}_{i\cdot\cdot} - \bar{\underline{Y}}_{\cdot\cdot\cdot})(\bar{\underline{Y}}_{i\cdot\cdot} - \bar{\underline{Y}}_{\cdot\cdot\cdot})'$$

$$\underline{S}_e(Y) = \sum_{i=1}^{r} \sum_{h(i)=1}^{n_i} (\bar{\underline{Y}}_{ih(i)\cdot} - \bar{\underline{Y}}_{i\cdot\cdot})(\bar{\underline{Y}}_{ih(i)\cdot} - \bar{\underline{Y}}_{i\cdot\cdot})'$$

$$(\underline{S}_h + \underline{S}_e)(Y) = \sum_{i=1}^{r} \sum_{h(i)=1}^{n_i} (\bar{\underline{Y}}_{ih(i)\cdot} - \bar{\underline{Y}}_{\cdot\cdot\cdot})(\bar{\underline{Y}}_{ih(i)\cdot} - \bar{\underline{Y}}_{\cdot\cdot\cdot})' \quad .$$

Zusätzlich ist im multivariaten Fall wieder von Interesse, welche der d abhängigen Variablen vornehmlich einen festgestellten Effekt der unabhängigen Variable widerspiegeln. Als Testverfahren eignet sich neben den stets verwendbaren konservativen Bonferroni-oder Bonferroni-Holm Tests das in Abschnitt 5.3.1 entwickelte und in Abbildung 5.1 dargestellte multiple sequentielle Testverfahren zum multiplen Niveau α, in dem die berechneten multiplen p-Werte für eine Testentscheidung gegen das für den Test auf einen Haupteffekt des Plotfaktors festgestellte α-Niveau verglichen werden. Beim Test auf einfache Plot-Haupteffekte ist es entsprechend α/c.

Für eventuell anzuschließende multivariate Vergleiche von Paaren von Plotstufen, gibt es für die zusätzlichen multiplen Tests bezüglich der d Dimensionen beim Bonferroni-, Bonferroni-Holm Test (mit/ohne Verbesserung nach Hommel (1985)) und beim Verfahren von Ryan (vgl. Abb. 5.2) keine Probleme. Die multiplen p-Werte sind jeweils gegen das fest vorgegebene reduzierte Fehlerniveau 1. Art zu vergleichen. In der komplizierteren Peritz-Prozedur (vgl. (5.40) ist solch ein festes Niveau nicht für alle Teilhypothesen-Entscheidungen gegeben. Für die 'contentious'-Hypothesen (vgl. (5.40) und Petrondas & Gabriel (1983)) wird ja aufgrund der Entscheidungen bzgl. anderer Teilhypothesen entschieden. Wenn man zu Testentscheidungen kommen will, gibt es dann nur die Möglichkeit, bei verworfenen 'contentious'-Hypothesen die multiplen p-Werte für die Dimensions-Teilhypothesen auch gegen das konservative Ryan- oder das etwas weniger konservative Welsch-Einot-Niveau zu vergleichen. Das Peritz-Verfahren für multivariate Beobachtungen ist so programmiert, daß man wählen kann, ob nur für die

globale Hypothese und die Paarvergleiche oder für alle signifikanten Teilhypothesen anschließend jeweils die multiple Testprozedur für die d abhängigen Variablen auszuführen ist (vgl. Abschnitt 8.4.5).

Beim Testen auf irgendeinen Einfluß des Split-Faktors in Verallgemeinerung von (7.4) bzw. (7.5) sind

$$\underline{S}_h^{B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c} n_i (\bar{Y}_{i.j}-\bar{Y}_{i..})(\bar{Y}_{i.j}-\bar{Y}_{i..})'$$

$$\underline{S}_e^{B,AB}(Y) = \sum_{i=1}^{r}\sum_{h(i)=1}^{n_i}\sum_{j=1}^{c} (Y_{ih(i)j}-\bar{Y}_{i.j})(Y_{ih(i)j}-\bar{Y}_{i.j})' \tag{7.13}$$

$$(\underline{S}_h+\underline{S}_e)^{B,AB}(Y) = \sum_{i=1}^{r}\sum_{h(i)=1}^{n_i}\sum_{j=1}^{c} (Y_{ih(i)j}-\bar{Y}_{i..})(Y_{ih(i)j}-\bar{Y}_{i..})'$$

unter den Intrablockpermutationen $\tilde{\Omega}_N = \times_{i=1}^{r} \times_{h(i)=1}^{n_i} \tilde{S}_c$ mit einer der MANOVA-analogen Teststatistiken zu analysieren.

Die Tests, welche pauschal das Vorhandensein von Effekten und ihren Wechselwirkungen analysieren, sind ebenfalls leicht auf den multivariaten Fall übertragbar. Für die Verallgemeinerung der Teststatistik aus (7.6) für den dCRF-rc benutzt man:

$$\underline{S}_h^{A,B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} (\bar{Y}_{ij.}-\bar{Y}_{...})(\bar{Y}_{ij.}-\bar{Y}_{...})' \tag{7.14}$$

$$\underline{S}_e^{A,B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} (Y_{ijk}-\bar{Y}_{ij.})(Y_{ijk}-\bar{Y}_{ij.})'$$

$$(\underline{S}_h+\underline{S}_e)^{A,B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} (Y_{ijk}-\bar{Y}_{...})(Y_{ijk}-\bar{Y}_{...})' \quad .$$

Die gewählte multivariate Teststatistik, z.B. die Pillai`sche Spur, wird unter allen $N!/\prod_i \prod_j n_{ij}!$ Permutationen von d-dimensionalen Beobachtungen zwischen den rc Zellen des Versuchsplanes bestimmt.

Im zweifaktoriellen Blockplan hat man statt (7.7) bzw. (7.8)

$$\underline{S}_h^{A,B,AB}(Y) = \sum_{i=1}^{r}\sum_{j=1}^{c} n (\bar{Y}_{.ij}-\bar{Y}_{...})(\bar{Y}_{.ij}-\bar{Y}_{...})' \tag{7.15}$$

$$\underline{S}_e^{A,B,AB}(Y) = \sum_{h=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{c} (Y_{hij}-\bar{Y}_{h..}-\bar{Y}_{.ij}+\bar{Y}_{...})(Y_{hij}-\bar{Y}_{h..}-\bar{Y}_{.ij}+\bar{Y}_{..}$$

$$(\underline{S}_h+\underline{S}_e)^{A,B,AB}(Y) = \sum_{h=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{c} (Y_{hij}-\bar{Y}_{h..}-\bar{Y}_{.ij}+\bar{Y}_{...})(Y_{hij}-\bar{Y}_{h..}-\bar{Y}_{.ij}+\bar{Y}_{..}$$

mit den unter der Nullhypothese zulassigen $((rc)!)^n$ Intrablock-Vektorpermutationen.

Die zuvor vorgestellten Testverfahren für dreifaktorielle Pläne und Abfolgen von Tests auf (einfache) Haupteffekte und Wechselwirkungen sowie mehrfache Vergleiche sind natürlich auch in ihrer multivariaten Form verwendbar. Die Einschränkungen für den Split-Plot-Plot Plan gelten analog. So komplexe multivariate Testprobleme durften aber in der experimentellen, wissenschaftliche Hypothesen testenden Forschung (vgl. Hager & Westermann, 1983c, S. 115f) selten vorkommen.

7.4 Verfahren zur möglichst strengen Prüfung wissenschaftlicher Hypothesen mit Signifikanztests

7.4.1 Überblick

Will man innerhalb eines größeren Forschungsprogrammes nicht mehr vornehmlich wissenschaftliche Hypothesen generieren, sondern diese einer möglichst strengen Prüfung (Popper, 1974, 1976 sowie Gadenne, 1976) unter Verwendung von Signifikanztests unterziehen, ergeben sich Konsequenzen sowohl für die Wahl geeigneter statistischer Verfahren wie insbesondere fur die Festsetzung adäquater maximaler Fehlerwahrscheinlichkeiten für die Entscheidung über die statistische(n) Hypothese(n). Häufiger sind die üblichen bekannten Teststrategien aus der statistischen Methodenliteratur nicht geeignet oder sogar im Widerspruch zu der Maxime, das Risiko einer falschen Entscheidung bezüglich der interessierenden wissenschaftlichen Hypothese (WH) zu minimieren.

Im Folgenden sei vorausgesetzt, daß eine WH sich sinnvoll dadurch überprüfen läßt, daß man eine oder mehrere statistische Hypothesen aus ihr ableitet (Bredenkamp, 1969, 1972, 1980; Hager & Westermann, 1983a, b, c; Meehl, 1967; Westermann, 1982; Westermann & Hager, 1982). Damit kann es aufgrund eines Experimentes (oder allgemeiner einer empirischen Untersuchung) bezüglich der WH zu zwei Typen von Fehlentscheidungen kommen (Hager & Westermann, 1983b):

e: Wahrscheinlichkeit für eine fälschliche Bewährung:
die WH ist nicht gültig, wird aber für gültig gehalten.

(7.16)

f: Wahrscheinlichkeit für eine fälschliche Nichtbewährung:
die WH ist gültig, wird aber für nicht gültig gehalten.

Im allgemeinen ist nur die genauere Kontrolle desjenigen Anteils an diesen beiden Wahrscheinlichkeiten möglich, der zu Lasten von falschen Entscheidungen bzgl. der/den aus der wissenschaftlichen Hypothese abgeleiteten statistischen Hypothese(n) SH ($SH_1,...,SH_n$) geht (Hager & Westermann, 1983b). Entsprechend definiert man folgende i.a. möglichst klein zu haltende Wahrscheinlichkeiten. Um die Vergleichbarkeit mit den zitierten Arbeiten zu erleichtern, werden die Bezeichnungen für die Wahrscheinlichkeiten nicht geändert:

ε : Wahrscheinlichkeit einer Entscheidung für die abgeleitete(n) SH(en) obwohl sie nicht gültig ist (sind)

(7.17)

ϕ : Wahrscheinlichkeit einer Entscheidung gegen die abgeleitete(n) SH(en) obwohl sie gültig ist (sind).

Da eine abgeleitete SH entweder der (einfachen oder zusammengesetzten) Nullhypothese oder der (einfachen oder zusammengesetzten) Alternative eines statistischen Tests entsprechen kann, sind im Fall einer SH ε und ϕ meist mit α und β bzw. β und α identisch, je nachdem, ob die SH einer Alternativhypothese oder einer Nullhypothese entspricht:

$$\text{SH} \Rightarrow H_1: \quad \varepsilon = \alpha, \quad \phi = \beta$$

(7.18)

$$\text{SH} \Rightarrow H_0: \quad \varepsilon = \beta, \quad \phi = \alpha \quad .$$

Häufiger ist der Fall, in dem die Ergebnisse der Prüfung mehrerer SHen zur Beurteilung einer WH erforderlich sind. Dann sind ε und ϕ nicht mehr mit α oder β zu identifizieren, sondern abhängig von den Fehlerwahrscheinlichkeiten 1. und 2. Art der n einzelnen statistischen Tests. In diesem Fall können oft maximale Fehlerwahrscheinlichkeiten α_M bzw. β_M für mindestens einen Fehler 1. Art oder 2. Art spezifiziert werden, die dann mit ε bzw. ϕ gemäß (7.18) identifiziert werden. Die Fehler 1. und 2. Art für die einzelnen Tests sind dann u. U. zu reduzieren, um α_M bzw. β_M nicht zu überschreiten. Wie in Abschnitt 7.4.2 an einigen typischen Fällen noch einmal demonstriert wird, hat die Strategie, die Fehler 1. oder 2. Art zu reduzieren, die aus der fälschlichen Verwerfung derjenigen SH(en) resultieren, die aus der WH abgeleitet wurden, u.U. andere Strategien als die übliche

α-Adjustierung für bestimmte Familien von Hypothesen zur Folge.

Wenn einerseits die WH als SH eine Konjunktion von Alternativhypothesen impliziert, ist eine α-Adjustierung für die einzelnen Tests nicht erforderlich, da eine falsche Entscheidung für die SH bereits dann vorliegt, wenn trotz Geltens mindestens einer der implizierten Nullhypothesen in diesem(n) Fall(Fällen) stets ein Fehler 1. Art begangen wird. Deshalb ist $\varepsilon=\alpha$ zu setzen, wenn alle Einzeltests zum selben nominellen Niveau durchgeführt werden, ansonsten ist $\varepsilon = \max_{i=1,..,n} \{\alpha_i\}$ geeignet. Die Fehler 2. Art sind pro Test zu reduzieren, da die SH bei Vorliegen von mindestens einem Fehler 2. Art zu Unrecht verworfen würde. Für die Wahrscheinlichkeit für mindestens einen Fehler 2. Art in n abhängigen oder unabhängigen Tests ist dann stets folgende Beziehung gültig (vgl. Abschnitt 5.2 für die Bezeichnung):

$$(7.19)\quad P_{\bigcap_{i=1}^{n} H_1^i}\left(\bigcup_{i\in J}\{\phi_i=0\}\right) \underset{(=)}{<} 1 - \prod_{i=1}^{n}(1-\beta_i) \quad \text{mit } J\subseteq\{1,..,n\} \text{ und } |J| \geq 1 \ .$$

Für identische $\beta_i \equiv \beta$ vereinfacht sich die Beziehung zu

$$\underset{(=)}{<} 1 - (1-\beta)^n \ .$$

Bei abhängigen Tests ist die obere Schranke i.a. eine gute Annäherung (Hager & Westermann, 1983a, Kimball, 1951; Miller, 1981, S. 101f; Petrinovich & Hardyck, 1969).

Bei Festlegung eines insgesamt maximal zulässigen $\phi=\beta_M$ werden die einzelnen Fehler 2. Art, falls sie alle gleichgroß sein sollen, zu:

$$(7.20)\quad \beta_i = 1 - (1-\beta_M)^{1/n} \ , \ i=1,..,n \ ,$$

oder als gute Näherung zu

$$\beta_i = \beta_M/n \ .$$

festgesetzt (Hager & Westermann, 1983a, c).

Probleme einer geeigneten Festsetzung der Effektstärke für diesen Fall werden von Bredenkamp (1984, S. 226f) diskutiert.

Wenn andererseits aus der WH als SH eine Konjunktion von Nullhypothesen folgt, ist eine analoge α-Adjustierung erforderlich mit $\phi=\alpha_M$, aber keine β-Reduktion pro Test, d.h. $\varepsilon=\beta_M=\beta$ oder allgemeiner $\varepsilon = \max_{i=1,..,n}\{\beta_i\}$;

$$(7.21)\quad P_{\bigcap_{i=1}^{n} H_0^i}\left(\bigcup_{i\in I}\{\phi_i=0\}\right) \leq 1-\prod_{i=1}^{n}(1-\alpha_i) \quad (=) \quad \text{mit } I\subseteq\{1,..,n\} \text{ und } |I|\geq 1 \ .$$

und $\alpha_i=1-(1-\alpha_M)^{1/m}, i=1,..,n$, oder $\alpha_i=\alpha_M/n$.

In einer dritten typischen Situation impliziert die WH als SH eine Konjunktion von n_1 Null- und n_2 Alternativhypothesen mit $n=n_1+n_2$. Westermann (1982) schlägt in solch einem Fall vor, wenn o.B.d.A. die ersten n_1 Hypothesen Alternativhypothesen und die restlichen n_2 Elemente Nullhypothesen sein sollen, die Fehlerwahrscheinlichkeit folgendermaßen festzusetzen:

$$(7.22)\quad \begin{aligned} \beta_1 = \ldots = \beta_{n_1} = \alpha_{n_1+1} = \ldots = \alpha_{n_1+n_2} = \phi/n \\ \alpha_1 = \ldots = \alpha_{n_1} = \beta_{n_1+1} = \ldots = \beta_{n_1+n_2} = \varepsilon/n \ . \end{aligned}$$

Allgemeiner könnte man auch zulassen, daß in (7.20) nur gilt: $\sum_{i=1}^{n}\beta_i = \beta_M = \phi$, in (7.21) $\sum_{i=1}^{n}\alpha_i = \alpha_M = \phi$

und in (7.22) $\sum_{i=1}^{n_1}\beta_i + \sum_{i=n_1+1}^{n}\alpha_i = \phi$ und $\sum_{i=1}^{n_1}\alpha_i + \sum_{i=n_1+1}^{n}\beta_i = \varepsilon$.

Diese obigen Empfehlungen stehen, wie deren Autoren auch feststellen, oft in verschiedener Hinsicht im Gegensatz zu der üblichen Anwendung statistischer Verfahren nicht nur in der experimentalpsychologischen Forschung:

(i) Die im Anschluß an einen signifikanten varianzanalytischen globalen Test eines Faktors ausgeführten multiplen Vergleiche beziehen sich häufig implizit auf eine viel größere Familie von Hypothesen als die von der WH implizierte (Konjunktion von) statistische(n) Hypothese(n) (vgl. Hager & Westermann, 1983, a, b. (Fall 8, S. 79), c).

(ii) Diese multiplen Testverfahren kontrollieren i.a. nur das globale oder bestenfalls das multiple Fehlerniveau 1. Art. Andererseits wird bei multiplen Tests, die das multiple α-Niveau einhalten, sichergestellt, daß die Wahrscheinlichkeit alle wahren Nullhypothesen nicht zu verwerfen, größer oder gleich 1-α ist. Inwieweit aber z.B. das Peritz-Verfahren aus Abschnitt 5.4.1 durch sein sequentielles Vorgehen und Verwerfen bzw. Akzeptieren von Teilhypothesen per Implikation nicht auch nur wenig an Teststärke z.B. für die i.a. allerdings zweiseitigen Elementarhypothesen verliert, wurde noch nicht untersucht.

(iii) Besonders der erste zuvor dargestellte Fall ohne α - Adjustierung aber mit β-Reduktion je Teilhypothese ist kaum beachtet worden. Die Konsequenz einer fehlenden oder nur partiellen Stützung auch 'fundierter' Hypothesen aufgrund der geringen Teststärke der einzelnen Tests ist offensichtlich. Ein Nachteil dieser Strategie ist allerdings häufig der sehr große erforderliche Stichprobenumfang schon pro Einzelhypothese der Testfamilie, wenn die minimalen aufzudeckenden Effekte nicht als groß anzunehmen sind. Man kann dann nur das β_M-Niveau und/oder α_M vergrößern oder auch noch mittlere Effektstärken als mit den Nullhypothesen verträglich ansehen.

(iv) Besonders ungeeignet ist das übliche Vorgehen aus (i), wenn die WH als SH eine streng monotone Ordnung aller k Faktorstufen impliziert $\mu_1 < \mu_2 < \ldots < \mu_k$ und nicht nur wie in den üblicherweise verwendeten Tests auf monotonen Trend die Alternative $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$ und $\mu_i < \mu_{i'}$ für mindestens ein Paar $i \neq i' \in \{1,..,k\}$. Ähnlich verhält es sich bei Vergleichen gegen eine Kontrollbedingung. Auf diese Probleme wird in Abschnitt 7.4.3 eingegangen.

(v) Für multivariate Verfahren ist die übliche Praxis einer Adjustierung des Fehlerniveaus 1. Art z.B. nach der Bonferroni-Prozedur ebenfalls problematisch, wenn die WH als SH eine Alternativhypothese impliziert, die für jede der betrachteten d abhängigen Variablen gelten soll. Hager & Westermann (1983c, S. 116) schlagen deshalb auch vor, multivariate Verfahren eher zur Hypothesengenerierung einzusetzen. Wieviel die exakte sequentielle multiple Testprozedur für Permutationstests bezüglich der abhängigen Variablen in einem uni- oder mehrfaktoriellen Versuchsplan aus Abschnitt 5.3 allerdings an Teststärke gegenüber d univariaten Verfahren einbüßt, wenn eine nicht zu geringe Korrelation zwischen den abhängigen Variablen besteht, ist ebenfalls bisher nicht untersucht.

In dem folgenden Abschnitt 7.4.2 soll zusammenfassend dargestellt werden, wie man mit Hilfe von Permutationstests für 'typische' Probleme bei solch einer veränderten Testpraxis vorgehen kann. Es werden dafür keine neuen Permutationstestverfahren notwendig sein. Vielmehr wird unter Berücksichtigung der prinzipiellen Schwierigkeit, für Permutationstests apriori die Teststärke festzulegen (vgl. Abschnitt 7.1 und 7.2) vornehmlich überlegt, ob sich darüberhinaus gegenüber parametrischen Verfahren Unterschiede in

der Vorgehensweise ergeben. An den zu testenden statistischen Hypothesen bezüglich Lageparametern andert sich nichts. Verändert, d.h. zum Teil abgeschwächt sind die in der (den) SH(en) implizit oder explizit enthaltenen statistischen Hilfshypothesen (Bredenkamp, 1980, 1984, S. 225) über unabhängige, identisch, oft normalverteilte Zufallsvariablen. Sie sind ersetzt durch Permutationsinvarianz-Forderungen bzgl. der Verteilung des gesamten Fehlervektors in Invarianz-Permutationstests bzw. durch (diskrete) Verteilungen von Design-Zufallsvariablen in Randomisierungstests, die ebenfalls Permutationsinvarianzeigenschaften besitzen. Nicht untersucht wird der Fall, in dem eine WH auf verschiedene Arten und unterschiedlich exakt oder vollständig formuliert werden kann und deshalb auch verschiedene statistische Hypothesen impliziert (Hager & Westermann, 1983c, Beispiel 5; 1984, S. 232ff; Bredenkamp, 1984).

7.4.2 Einige typische Fragestellungen

Um einen recht großen Bereich von typischen SHen als Folgerungen aus WHen zu erfassen, werden mehrere univariate einfaktorielle Probleme betrachtet. Einleitend sei noch einmal hervorgehoben, daß bei Verwendung von Randomisierungs- oder Permutationstests zu einer apriori-Festlegung des jeweils benötigten Mindeststichprobenumfanges ohne die in Abschnitt 6.3 vorgeschlagenen aufwendigen Nährungsverfahren nur die Verwendung der Resultate für den Fall mit (multivariater) Normalverteilung übrigbleibt. Rechtfertigen

ließ sich diese Vorgehensweise nur mit der asymptotischen Äquivalenz von Permutationstests unter Normalverteilungsbedingungen.

(1) Die WH impliziert eine einseitige Alternative im (un)verbundenen Zweistichprobenproblem

Dieser Fall wurde unter Verwendung des t-Test schon von Bredenkamp (1969, 1972) behandelt (vgl. z.B. auch Cohen, 1969; Guenther, 1977, 1981):

$$WH \Rightarrow SH: \quad H_1: \mu_1 > \mu_2 \quad (H_o: \mu_1 \leq \mu_2) \quad .$$

Der benötigte Stichprobenumfang je Stichprobe im unverbundenen Zweistichprobenproblem ist bei Verwendung der Normalverteilungsapproximation unter Festlegung von α, β und minimalem Effekt ω^2 gegeben durch:

(7.23) $$n = (z_\alpha + z_\beta)^2 \cdot (1-\omega^2)/2\omega^2 \quad .$$

ω^2 ist der Mindestanteil der gesamten Datenvarianz, der durch den Unterschied der beiden experimentellen Bedingungen zustande kommen soll; z_α, z_β sind die entsprechenden $(1-\alpha)$ bzw. $(1-\beta)$- Quantile der Standardnormalverteilung.
Diese approximative Bestimmung von n mit Hilfe der Standardnormalverteilung liefert i.a. hinreichend genaue Werte. Exaktere Bestimmungen von n mit Hilfe der (nichtzentralen) t-Verteilung (vgl. z.B. Johnson & Kotz, 1970; Bredenkamp 1972, S. 92) sind rechnerisch aufwendig und führen zu leicht größeren Stichprobenumfängen um 1 oder 2 Beobachtungseinheiten je Stichprobe. Ansonsten gibt es auch die Möglichkeit, die Tabellen von Cohen (1977) zu verwenden.

Für einen einseitigen Permutationstest kann man damit für dieses minimale n die Teststatistik (3.38) wählen. Nur bei großen Effekten und nicht zu kleinen Fehlern 1. und 2. Art wird n so klein ausfallen, daß eine exakte Ausführung des Permutationstests für alle unter der Nullhypothese zulässigen Permutationen möglich ist. Falls man eine approximative sog. Monte-Carlo Lösung mit einer großen Zufallsstichprobe aller zulässigen Permutationen wählt oder wegen dem sonst entstehenden Rechenaufwand wählen muß (vgl. Abschnitt 8.2), ist noch eine Korrektur für das minimale n erforderlich. Um die gleiche Teststärke wie für den exakten Permutationstest zu erhalten, muß das n aus (7.23) durch die sog. Dwass-Effizienz (vgl. Tabelle 8.3), die stets kleiner als 1 ist, dividiert werden. Dieser Vorschlag wird in Abschnitt 8.2 näher begründet und stellt eine Korrektur wie bei der Verwendung von nichtparametrischen Rangtests statt der analogen Tests mit Normalverteilungsannahme dar. Dort wird, einem Vorschlag von Lienert (1973) folgend (vgl. Bredenkamp, 1980, S. 62; Hager & Westermann, 1983a, S. 184) durch die asymptotische Effizienz des Rangverfahrens zum Test mit Normalverteilungsannahme im Normalverteilungsfall dividiert.

Nach Ausführung des Permutationstests kann man etwa nach folgender, drei Situationen unterscheidender Entscheidungsstrategie (vgl. Bredenkamp, 1972, S. 80ff; Hager & Westermann, 1983a, S. 188ff mit leicht unterschiedlichen Akzenten) folgen:

(i) Der p-Wert ist größer als α; die Nullhypothese wird akzeptiert. Es wird also eine Entscheidung zugunsten der Nullhypothese vorgenommen! Die WH wird vorerst als nicht bewährt angesehen.

(7.24) (ii) Der p-Wert ist kleiner als α und der aus den Daten geschätzte experimentelle Effekt ist nicht kleiner als der zuvor festgelegte Mindesteffekt. Die WH wird als (vorerst) bewahrt angesehen.

(iii) Der p-Wert ist kleiner oder gleich α aber der geschätzte experimentelle Effekt ist kleiner als der festgelegte Mindesteffekt. Es liegt (schwache) Evidenz für die WH vor. Replikationen mit kleinerem Mindesteffekt und damit größerem Mindeststichprobenumfang sind sinnvoll. (Siehe aber auch Hager & Westermann, 1983a, S. 191 für den zweiseitigen Fall).

Auch der vierte, bei Hager & Westermann (1983a, S. 189) angeführte Fall von in Replikationen aufzufindenden konsistenten kleinen Effekten in die erwartete Richtung der WH allerdings mit p-Werten größer oder gleich α sei erwähnt (vgl. auch Bredenkamp, 1969, S. 283).

Als (Punkt-) Schätzung des experimentellen Effektes aus den Daten des Experimentes hat man:

$$(7.25) \quad \hat{\omega}^2 = (\bar{y}_{1.}-\bar{y}_{2.})^2/(4\cdot q^2_{Fehler}+(\bar{y}_{1.}-\bar{y}_{2.})^2)$$

mit $$q^2_{Fehler} = 1/(2(n-1))\left(\sum_{j=1}^{n}(y_{1j}-\bar{y}_{1.})^2+\sum_{j=1}^{n}(y_{2j}-\bar{y}_{2.})^2\right)$$

(vgl. Scheffe, 1959, S. 339).

Wie gut diese nicht erwartungstreue Punktschätzung des Parameters ω^2 allerdings ist, ist unbekannt (vgl. Carroll & Nordholm, 1975; Lane & Dunlap, 1978; Hager & Westermann, 1983a, Kap. 9.3). Unter Umständen sollte man auch überlegen, ob nichtparametrische Schätzer für $\mu_1-\mu_2$, falls keine Normalverteilung vorliegt, nicht geeigneter und robuster

sind. Lehmann (1975, S. 82ff) schlägt z.B. vor, den Median aller n^2 Differenzen $Y_{1j}-Y_{2j'}$, $-j,j'=1,..,n$ - einen sog. Hodges-Lehmann Schätzer (vgl. z.B. Randles & Wolfe, 1979, S. 209) zu betrachten. In diesem Zusammenhang soll auch noch einmal die Bemerkung von Scheffe erwähnt werden, daß die Schätzung von σ^2_{Fehler} anfällig ist gegen Verletzungen der Normalverteiltheit.

Deshalb erscheint es für Permutationstests sinnvoll, zusätzlich die naive bedingte Teststärke zum erreichbaren Niveau α nach Ausführung des Experimentes wie in Abschnitt 6.3. (Formel 6.29)) beschrieben, zu schätzen. Ist diese geschätzte Teststärke dann kleiner als die der Berechnung des minimalen Stichprobenumfanges zufgrunde gelegte (apriori) Teststärke, liegt wie im Fall (iii) in (7.24) nur schwache Evidenz für die WH vor.

Guenther (1977, 1981, 1982) schlägt vor, μ_2 als p-Quantil mit $p<0.5$ der Verteilung mit Erwartungswert μ_1 zu spezifizieren. Dann erhält man approximativ als minimales n:

(7.26)
$$n = 2\cdot(z_\alpha+z_\beta)^2/z_p^2 + 1/4\cdot z_\alpha^2 \quad .$$

Im verbundenen Zweistichprobenproblem hat man als minimalen Stichprobenumfang im Normalverteilungsfall, wenn man die Differenz der Erwartungswerte $\delta=(\mu_1-\mu_2)$ als Vielfaches von $\sigma(Y_1-Y_2)$ definiert:

(7.27) $$n = (z_\alpha+z_\beta)^2/V^2$$

mit $V=\delta/\sigma(Y_1-Y_2)$ (vgl.Bredenkamp, 1969, S. 280).

Im verbundenen Zweistichprobenproblem ist unter sonst gleichen Bedingungen n aus (7.27) nur halb so groß wie im unverbundenen Zweistichprobenproblem.

Als Teststatistik für den Permutationstest wählt man die zweite in (3.39) angeführte Teststatistik für den einseitigen Fall und verfährt wie zuvor für den Fall des unverbundenen Problems beschrieben.

Die Effektschätzung erhält man mit folgender Formel:

(7.28) $\hat{V}^2 = (\bar{y}_{1.}-\bar{y}_{2.})/(1/(n-1)\sum_{i=1}^{n}(y_{1i}-y_{2i}-(\bar{y}_{1.}-\bar{y}_{2.}))^2)$.

Ein Hodges-Lehmann Schätzer für δ ist der Median aller $(D_i+D_{i'})/2$ mit $i\leq i'=1,..,n$ und $D_i=Y_{1i}-Y_{2i}$ (Randles & Wolfe, 1979, S. 204).

Guenther (1981, 1982) zeigt, wenn μ_2 als p-Quantil der Verteilung mit μ_1 angegeben wird mit $p<0.5$, daß n sich berechnen läßt aus:

(7.29) $n = (z_\alpha+z_\beta)^2/z_p^2 + 0.5z_\alpha^2$.

Der Fall, daß die WH eine zweiseitige Alternative impliziert, ist in der theorietestenden psychologischen Forschung selten. Die Formeln (7.23) und (7.27) können, wenn man dort α durch $\alpha/2$ ersetzt, weiter benutzt werden. Bei Verwendung des Permutationstests ist die für den zweiseitigen Fall geeignete Teststatistik (3.36) bzw. (3.39) zu wählen.

Falls die WH die Nullhypothese des (un)verbundenen Zweistichprobenproblems impliziert, ist ebenfalls zweiseitig zu testen. Die Entscheidungsstrategie aus (7.24) ist sinngemäß so zu ändern, daß im Fall (i) sich die WH bewährt hat, nicht aber im Fall (ii) und daß im Fall (iii) schwache Evidenz gegen die WH vorliegt.

(2) Die WH impliziert eine Abweichung von der Nullhypothese im (un)verbundenen k Stichproben-problem

Die in der statistischen Literatur über multiple Tests üblicherweise betrachtete Alternative, die vornehmlich auch in Abschnitt 5 behandelt wurde, nimmt lediglich an, daß zwischen mindestens zwei der insgesamt k Lageparameter irgendein unspezifizierter, meist ungerichteter Unterschied besteht. Die zugehorige WH dürfte, wie Bredenkamp (1980, S. 56) und Hager & Westermann (1983b) argumentierten, in Experimenten in der Psychologie, die eine psychologische Hypothese testen, nicht von Interesse sein. Für explorative Zwecke ohne spezifische Vorhersagen und Erklärungen haben die in Abschnitt 4.5 dargestellten Verfahren dagegen ihre Berechtigung. Spezifische Alternativen, die durch die jeweilige(n) WH(en) impliziert werden, beziehen sich üblicherweise auf eine der folgenden Situationen:

(i) mehrere orthogonale Kontraste
(ii) mehrere nicht-orthogonale Kontraste
(iii) multiple t-Tests
(iv) Vergleich(e) mit einer Kontrollbedingung
(v) teilweiser oder vollständiger monotoner Trend in den Faktorstufenerwartungswerten
(vi) lineare oder allgemeinere polynomiale Trends.

Auch wenn aus der WH als SH ein von Null verschiedener Kontrast oder Trend folgt, kann es sinnvoll sein, vorab mit einem globalen Permutationstest fur alle k Faktorstufen mit der Teststatistik (3.40) im unverbundenen Fall oder mit (3.41) im verbundenen Fall zu prüfen, ob überhaupt Unterschiede zwischen den Faktorstufen vorliegen (Bredenkamp,

1980, S. 56; Hager & Westermann, 1983a, S. 124). Allerdings sollten α und β wie auch der minimal aufzudeckende Effekt nicht zu groß angenommen werden.

Der im unverbundenen k-Stichprobenproblem minimal benötigte Gesamtstichprobenumfang im ANOVA-Fall berechnet sich approximativ aus (vgl. Bredenkamp, 1980, S. 53; Cohen, 1977, S. 439; Hager & Möller, 1983):

(7.30) $N = L/f^2 + k$

mit dem vorgegebenen Nichtzentralitätsparameter

$$f^2 = \sum_{i=1}^{k} (\mu_i-\mu)^2/(k\cdot\sigma^2_{Fehler}) \quad (= \lambda/N)$$

und L, dem von Cohen (1977) und Hager & Möller (1983) für verschiedene α, β und $k=df_{Fehler}+1$ vertafelten approximativ bestimmten Nichtzentralitätsparameter der F-Verteilung.

Wie z.B. Bredenkamp (1980, Anhang) erläutert, läßt sich der Parameter f^2 darstellen als:

(7.31) $f^2=R^2/(1-R^2)$

mit R^2 dem Quadrat der multiplen Korrelation zwischen der in k-1 orthogonale Prädiktoren kodierten unabhängigen Variablen und der abhängigen Variablen. Diese Definition ist auch auf mehrere unabhängige Variablen erweiterbar, was im Folgenden noch benutzt wird. Im CR-k gilt außerdem $R^2=SS_{Faktor}/SS_{Total}$.

Mit der Spezifizierung von R^2 kann N bestimmt werden. Tatsächlich wählt man die Zahl $N^*\geq N$, die ohne Rest durch k teilbar ist, sodaß als Stichprobenumfang je Faktorstufe N^*/k resultiert. Bei kleineren α und β wird in aller Regel nur die Monte-Carlo Lösung des Permutationstests anwendbar sein, sodaß, wie zuvor erläutert, N^* noch durch die Dwass-Effizienz dividiert werden sollte. Außerdem sollte dieses

neue N^{**} noch 'ein wenig' vergrößert werden, da der Permutationstest nur asymptotisch im Normalverteilungsfall dieselbe Teststärke wie der ANOVA-Test besitzt, im finiten Fall aber häufig eine leicht niedrigere finite Teststärke zeigt (vgl. z.B. Baker & Collier, 1966b; Collier & Baker, 1966). Nach Ausführung des Permutationstests für die globale Nullhypothese braucht nicht weiter getestet zu werden, falls der p-Wert größer als α ausfällt. Die Schätzung von f^2 nach Ausführung des Experimentes kann mit Hilfe des Wertes der ANOVA F-Statistik erfolgen über (vgl. Bredenkamp, 1970):

(7.32) $\hat{f}^2 = \hat{R}^2/(1-\hat{R}^2)$ und $\hat{R}^2 = F/(F+(N-k)/(k-1))$

bzw. dem korrigierten $\hat{R}^{*2} = 1-(1-\hat{R}^2)(N-1)/(N-k)$.

Einsetzen von $\hat{R}^{*2}$ in (7.31) ergibt:

$\hat{f}^{*2} = (k-1)(F-1)/(N-1)$.

In theorietestenden Experimenten ist der bisher beschriebene Fall des Testens gegen eine beliebige Abweichung von der globalen Nullhypothese jedoch kaum gegeben. Deshalb sollen nun Permutationstests für die oben angeführten Alternativen (i) - (vi) vorgestellt werden.

Im Fall des parametrischen verbundenen k-Stichprobenproblems bestimmt man die minimal benötigte Anzahl n von Blöcken, nach den von Bredenkamp (1980, S. 92) angegebenen Formeln bei zufälligen Blöcken:

(7.33) $n = L/(f^2(k-1)) + 1$.

sodaß N = nk Beobachtungen resultieren. L ist aus den Tabellen von Cohen (1977, Kap. 9) für verschiedene α und β zu entnehmen und f^2 schätzt man mit einem ebenfalls bei Bredenkamp (1980, S. 72, 92) beschriebenen Kodierungstrick (siehe auch Pedhazur, 1977).

(i,ii) (Nicht)orthogonale Kontraste

Im Gegensatz zum parametrischen Fall sind bei der Verwendung von Permutationstests die Teststatistiken des parametrischen Falles nicht sinnvoll verwendbar, da im Nenner der t-Statistiken für die verschiedenen Kontraste stets derselbe Fehlerterm gewonnen aus allen k Bedingungen steht (vgl. Kirk, 1968, Kap. 3.2). Wie in Abschnitt 5.4 bereits argumentiert, bezieht sich die Nullhypothese eines Permutationstests jedoch stets nur auf die Vertauschbarkeit von Beobachtungen unter denjenigen Faktorstufen, die von der jeweiligen (Teil-)Hypothese berücksichtigt werden. Damit ist es auch konzeptuell problematisch, wie im parametrischen orthogonalen Fall den aufzudeckenden Gesamteffekt als Summe der Einzeleffekte für die verschiedenen (max. k-1) orthogonalen Kontraste festzulegen. Auch dort mußte zudem bei der Spezifizierung der Mindesteffekte je Kontrast beachtet werden, daß nicht ein insgesamt zu großer Mindesteffekt resultiert (vgl. Bredenkamp, 1984, S. 226ff).

Nur wenn man sich trotz dieser obigen Unterschiede im Modell für die Planung des Stichprobenumfangs am Normalverteilungsfall orientiert, sind die Vorschläge von Bredenkamp (1984, S. 227) verwendbar. Dort wird für $m \leq k-1$ lineare Kontraste der Form

$$(7.34)\quad \psi = \sum_{i=1}^{k} c_i \mu_i \qquad \text{mit } \sum_{i=1}^{k} c_i = 0\ ,\ \sum_{i=1}^{k} |c_i| = 2,\ I=\{c_i \mid c_i \neq 0\}\ ,\ |I| \geq 2\ .$$

der je Faktorstufe benötigte Stichprobenumfang n folgendermaßen berechnet:

(7.35) $n = (z_{\alpha}+z_{\beta/m})^2/(k\cdot f_{\psi}^2)$

mit $f_{\psi}^2=((\sum_{i\in I} c_i\mu_i)^2/k\cdot\sum_{i\in I} c_i^2)/\sigma^2_{Fehler}$ und z_{α}, $z_{\beta/m}$ den entsprechenden Quantilen der Standardnormalverteilung.

Als Teststatistik eines Permutationstests für das Vorliegen eines linearen Kontrastes wird

(7.36) $S_{\psi}(Y) = \sum_{i\in I} c_i\bar{Y}_{i.}$

verwendet. Die Permutationsverteilung wird unter allen $(n|I|)!/|I|n!$ zulässigen Permutationen der Beobachtungen zwischen den im linearen Kontrast berücksichtigten Faktorstufen bestimmt. Je nach Richtung der einseitigen Alternative $\psi\geq 0$ oder $\psi\leq 0$ ist der p-Wert als $p^{\geq}$ oder $p^{\leq}$ zu bilden. Um den im Experiment tatsächlich aufgetretenen Effekt zu schätzen, verwendet man je Kontrast

(7.37) $\hat{f}_{\psi}^2 = ((\sum_{i\in I} c_i\bar{y}_{i.})^2/k\sum_{i\in I} c_i^2)/\sigma^2_{Fehler}$.

Für die Fehlerschätzung kann man sich streng genommen nur auf die gepoolte Binnenvarianz der vom Kontrast betroffenen Faktorstufen

$$\hat{\sigma}^2_{\psi,Fehler} = 1/(|I|(n-1))\cdot\sum_{i\in I}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

stützen, die allein die Information aus den tatsächlich permutierten Werten nutzt. Im Normalverteilungsfall würde stattdessen für alle m Kontraste dieselbe Schätzung der Fehlervarianz über alle k Faktorstufen Verwendung finden. Gordon (1985) hat gezeigt, daß bei k-1 orthogonalen Kontrasten jeweils getestet mit $\alpha_c=\alpha$ im Anschluß an einen bei α signifikanten globalen F-Test es zu einem experimentweisen $\alpha_{EW}>\alpha$ kommen kann.

Falls es sich nicht um m=k-1 orthogonale Kontraste, sondern z.B. um m (abhängige) Vergleiche von je zwei Faktorstufen i, i' handelt, vereinfacht sich in (7.35) f_{ψ}^2 zu

$$f_{i,i'}^2 = (\mu_i - \mu_{i'})^2 / (k \cdot 2 \cdot \sigma_{i,i',Fehler}^2) \quad .$$

Die Teststatistik aus (7.36) hat dieselbe Gestalt wie in Formel (3.38) für das unverbundene Zweistichprobenproblem. Wie Bredenkamp (1984, S. 227) dargestellt hat, addieren sich auch im parametrischen Fall die Effekte nicht zum Gesamteffekt (7.32) für alle k Faktorstufen.
Im verbundenen k-Stichprobenproblem kann man zum Testen (orthogonaler) linearer Kontraste ebenfalls die Teststatistik aus Formel (7.36) ohne vorherige Eliminierung der Blockeffekte verwenden. Für die Permutationsverteilung von $S_{\psi}(Y)$ sind allerdings nur die $(n!)^{|I|}$ Intrablockpermutationen über die im Kontrast berücksichtigten $|I|$ Faktorstufen auszuführen. Je nach Richtung der einseitigen Alternative ist wiederum der p-Wert als $p^{\geq}$ oder $p^{\leq}$ zu bilden.

(iii) Multiple_t-Tests

Gegenüber den Testfamilien aus Abschnitt 5 und auch gegenüber den zuvor dargestellten Tests für orthogonale Kontraste haben besonders Hager & Westermann (1983a, S. 125f) die Vorzüge multipler t-Tests betont, besonders wenn etwa nicht alle möglichen paarweisen Vergleiche im (un)verbundenen k-Stichprobenproblem durchgeführt werden oder wenn die WH nur SHen bezüglich einer Teilmenge dieser Paarvergleiche impliziert. Bei Verwendung von Permutationstests für das (un)verbundene Zweistichprobenproblem ergibt sich ein weiterer Vorteil für diese Vorgehensweise.

Die erforderlichen Fehlerschätzungen bleiben auch konzeptuell bei der Verwendung von multiplen Zweistichprobentests auf die von der jeweiligen Hypothese betroffenen Faktorstufen beschränkt. Selbstverständlich muß zur Planung des erforderlichen Stichprobenumfanges mit den in Abschnitt 7.4.2(1) beschriebenen Methoden wieder die asymptotische Äquivalenz der Teststatistik zum parametrischen t-Test für die approximative Stichprobenumfangsbestimmung benutzt werden. Auf die notwendige Reduktion des β-Fehlers je Vergleich zweier Stichproben sei noch einmal hingewiesen.

(iv) Vergleiche mit einer Kontrollbedingung

Die Elementarhypothesenfamilie aus Abschnitt 5.4.2 für k-1 Vergleiche je einer von k-1 Bedingung gegen ein und dieselbe Kontrollbedingung führt zu speziellen (nicht-orthogonalen) linearen Kontrasten. Die zuvor dargestellte multiple sequentielle Testprozedur testet aber nicht nur Elementarhypothesen, so daß eine Testplanung mit α- und β- Fehlerkontrolle sowie Spezifizierung eines Mindesteffektes je Vergleich sich auf andere Testverfahren stützen muß. Bredenkamp (1980, S. 60f) hat ein Verfahren für den parametrischen Fall vorgeschlagen, in dem t-verteilte Teststatistiken der Form

(7.38) $S_i(Y) = (\bar{Y}_{i.} - \bar{Y}_{k.}) / (2 \cdot SS_{total}(1-\hat{R}^2)/N(n-1))^{1/2}$, i=1,..,k-1,

mit $\hat{R}$ wie in (7.32) und mit N-k Freiheitsgraden verwendet werden.

Beim Vergleich einer experimentellen Bedingung mit der Kontrollbedingung wird in dieser Teststatistik wieder auf eine Fehlerschätzfunktion über alle Bedingungen hinweg zurückgegriffen. Für die Planung von Permutationstests ist deshalb die approximative Teststärkebestimmung mit Hilfe von

(un)verbundenen einseitigen t-Tests, für die bei jedem der k-1 Vergleiche gegen die Kontrollbedingung auch unterschiedliche Mindesteffekte angenommen werden können, geeignet. Mit den in 7.4.2.(1) dargestellten Verfahren werden der Stichprobenumfang und der experimentelle Effekt geschätzt. Falls für alle k-1 einseitigen Vergleiche identische Mindesteffekte mit gleicher Teststärke angenommen werden, ist dort lediglich z_{β} durch $z_{\beta/(k-1)}$ zu ersetzen. Bei unterschiedlichen Mindesteffekten resultieren je Vergleich unter Umständen unterschiedliche Stichprobenumfänge für die Kontrollbedingung. Dann wird man für die Kontrollbedingung das Maximum dieser Stichprobenumfänge wählen (Bredenkamp, 1980, S. 56f).

(v) Lineare allgemeine polynomiale und monotone Trends

Da diese Arbeit auf das Testen linearer Hypothesen in linearen Modellen bezüglich qualitativ abgestufter Ausprägungen der unabhängigen Variablen beschränkt ist, wurden bisher keine Tests auf lineare oder allgemeine polynomiale Trends auf ihre Übertragbarkeit auf den Permutationsfall hin analysiert. Wie z.B. Edgington (1980, Kap. 9) darstellt, lassen sich für eine Vielzahl von Fragestellungen Trendtests entwickeln. Er schlägt einerseits Tests der üblichen Nullhypothesen im (un)verbundenen k-Stichprobenproblem vor (Kap. 9.1-9.6), die bei einer speziellen Trendalternative geeignet sind und deshalb von ihm als 'goodness-of-fit' Trendtests bezeichnet werden. Im unverbundenen k-Stichprobenproblem mit k quantitativ abgestuften Faktorstufen ist eine generelle Form solch einer Teststatistik

$$(7.39)\quad S(Y) = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \hat{Y}_i)^2 \quad ,$$

wobei $\hat{Y}_i$ den durch den speziell interessierenden oder hypothetisch angenommenen Trend vorhergesagte Mittelwert für die n_i Beobachtungen für die i-te Stufe der quantitativen unabhängigen Variablen darstellt (vgl. Kirk, 1968, S. 124f). Überzufällige Abweichungen von dem speziell betrachteten Trend werden anhand der über alle $n_.!/\Pi n_i!$ zulässigen Permutationen zwischen Stufen der unabhängigen Variablen ermittelten Permutationsverteilung von (7.39) bestimmt. Als p-Wert ist $p^{\leq}$ zu wählen, da Abweichungen vom vorhergesagten Trend zu größeren Differenzen zu den Faktorstufenmitteln führen.

Als zweite Form von Trendtests behandelt Edgington (Kap. 9.7-9.11) 'correlation'-Trendtests, für die man unter dem Permutationsmodell eine zum Zähler der parametrischen F-Statistik für orthogonale polynomiale Trendkomponenten äquivalente Teststatistik benutzt:

$$(7.40)\quad S(Y) = \left(\sum_{i=1}^{k} c_i Y_{i.}\right)^2 \quad .$$

Dabei sind die c_i die entsprechenden, z.B. in Kirk (1980, Tab. D12) für gleichabständige quantitative Faktorstufen und gleiche Stichprobenumfänge vertafelten, orthogonalen Polynomialkoeffizienten. Für nicht gleichabständige Faktorstufen und/oder ungleiche n_i ist dort im Anhang C ein Algorithmus zur Berechnung der c_i angegeben.

Falls die Richtung des jeweiligen Trends und nicht nur der Grad des Polynoms von Interesse ist, muß das Quadrat in (7.40) entfallen und je nach Richtung $p^{\geq}$ oder $p^{\leq}$ bestimmt werden. Zulässige Permutationen bei Gültigkeit der globalen

Nullhypothese sind wie bisher alle $n_{\cdot}!/\prod_i n_i!$ Vertauschungen von Beobachtungen zwischen Faktorstufen im unverbundenen k-Stichprobenproblem. Im verbundenen k-Stichprobenproblem ist ebenfalls die Teststatistik (7.40) verwendbar, die unter den $(k!)^n$ Intrablockpermutationen betrachtet werden muß. $Y_{i.}$ ist nur durch $Y_{.j}$, j=1,..,k,zu ersetzen.

Auch der Test auf Abweichungen von einem z.B. linearen Trend ist leicht auf den Permutationstestfall zu übertragen, indem man

$$(7.41) \quad S(Y) = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 - \left(\sum_{i=1}^{k} c_{i,lin} \bar{Y}_{i.}\right)^2 / \sum_{i=1}^{k} n_i c_{i,lin}^2$$

betrachtet und in diesem Fall für einen linearen Trend geeignete c_i wählt. Zur Testentscheidung ist $p^{\geq}$ zu bilden unter den zuvor beschriebenen zulässigen Permutationen. Ebenfalls unproblematisch ist der globale Test auf das Verschwinden von Trendkomponenten höherer, z.B. k_o-ter Ordnung. Dann ist in (7.41) statt des Ausdrucks für die lineare Trendkomponente der Anteil für alle Komponenten 1. bis (k_o-1)-ter Ordnung zu subtrahieren.

Mit den zu Anfang dieses Abschnittes genannten grundsätzlichen Einschränkungen kann auch für Permutationstests eine rationale Planung des minimal erforderlichen Stichprobenumfangs, wie z.B. bei Bredenkamp (1980, S. 57ff) erläutert, erfolgen.

Die Analyse monotoner, nicht in ihrer funktionalen Form spezifizierter Trends wird in Abschnitt 7.4.3 dargestellt.

(3) Die WH impliziert eine oder mehrere (spezielle) Alternativen in einem mehrfaktoriellen Versuchsplan

Wegen der moglichen Vielzahl von aus einer WH abzuleitenden Kombinationen von SHen werden nur einige Beispiele von SHen betrachtet, die bereits in der psychologischen Methodenliteratur für den parametrischen Fall erörtert wurden. Zu beachten ist nachfolgend wieder, daß im Gegensatz zur parametrischen ANOVA mehrere Tests auf Haupt- und Wechselwirkungen in einem mehrfaktoriellen Versuchsplan nicht in einem Berechnungsgang erfolgen, sondern i.a. separate Tests mit unterschiedlichen Gruppen von unter der jeweiligen Nullhypothese zulässigen Permutationen auszuführen sind. Als Standardbeispiel dient häufig der zweifaktorielle vollständige Zufallsplan. Da es für diesen keinen Invarianz-Permutationstest auf Wechselwirkungen gibt, sind die Vorschläge von Bredenkamp (1980, S. 62ff) für den parametrischen Fall nicht durchgängig übertragbar.

Der Mindeststichprobenumfang für den globalen Test auf irgendeinen Effekt im zweifaktoriellen vollständigen Zufallsplan mit gleichen Stichprobenumfängen unter den rc Faktorstufenkombinationen läßt sich auch für den Permutationstest approximativ folgendermaßen abschatzen:

(7.42) $N = L/f^2 + rc$

mit $f^2 = R^2_{A,B,AB}/(1-R^2_{A,B,AB})$ und $R^2_{A,B,AB} = R^2_A + R^2_B + R^2_{AB}$.

Über obige Beziehung sind die Beiträge der Haupt- und Wechselwirkungen einzeln spezifizierbar. L ist wieder für vorgegebenes α, β fur (u=)rc-1 Freiheitsgrade bei Cohen (1977, Tab. 9.4.2) abzulesen.

Dieser Test erfolgt wie der Permutationstest im unverbundenen k=rc-Stichprobenproblem. Zur Effektschätzung nach Ausführung des Experimentes ist deshalb (7.32) geeignet, wenn man dort k nur durch rc ersetzt.

Besonders wenn einer der beiden Faktoren, etwa B, eine Kontrollvariable ist und/oder die experimentelle Fragestellung besonders auf den Faktor A bezogen ist, kann man mit einem Permutationstest nun noch prüfen, ob der Faktor A bzw. der Kontrollfaktor überhaupt einen Einfluß hat (vgl. Abschnitt 3.1.5(x)). Vor Ausführung dieses Tests mit der Teststatistik (3.60) unter unabhängiger Vertauschung der Werte innerhalb jeder Stufe des Faktors B (bzw. von A) ist jedoch abzuschatzen, ob der für die obige globale Nullhypothese minimal notwendige Umfang auch für diese Hypothese einen Test mit hinreichender Teststärke zuläßt. Wählt man beispielsweise die Effektstärke wie in Bredenkamp (1980, S. 65):

$$R^2_A = 0.15 \ , \ R^2_B = 0.05 \ , \ R^2_{A,B} = 0.05 \ ,$$

so waren dort für den parametrischen Fall und $\alpha=\beta= 0.05$ bei r=c=3 Faktorstufen aufgerundet 81/rc = 9 Beobachtungseinheiten pro Faktorstufenkombination erforderlich. Damit ergibt sich für

$$L_{A,AB}=(N-rc)f^2_{A,AB}=(81-9)(R^2_{A,B,AB}-R^2_B)/(1-R^2_{A,B,AB}) = 19.2 \ ,$$

und aus Tab. 9.3.2 in Cohen (1977) entnimmt man für $\alpha=0.05$ und $u=(r-1)(c-1)+(c-1)=c(r-1)$ eine Teststärke von ca. 0.93. Für die Feststellung irgendeines Effektes des Kontrollfaktors hätte man nur eine Teststärke von etwa 0.59. Bei Ausführung beider Tests ist noch zu beachten, daß auch im Fall des Permutationstests die beiden Teststatistiken nicht unabhängig sind und man mit einem reduzierten β-Niveau testen müßte.

Für die Analyse einfacher Haupteffekte muß bei Verwendung von Permutationstests strenggenommen ein Modellwechsel erfolgen, da die zulässigen Permutationen jeweils innerhalb einer Stufe des jeweils anderen zweiten Faktors auszuführen sind (vgl. Willmes & Pyhel, 1981 für den SPC-r.c Plan und Abschnitt 3.1.5). Dann müßte je einfachem Haupteffekt wie für den einfaktoriellen Zufallsplan mit reduziertem β-Niveau bei multiplen Tests auf einfache Haupteffekte verfahren werden. Falls Hypothesen bezüglich einfacher Haupteffekte für beide Effekte mit unterschiedlicher Effektstärke vorliegen, sollte stets der jeweils größere Stichprobenumfang gewählt werden. Falls alle einfachen Haupteffekte getestet werden sollen, kann das unter Umständen zur Folge haben, daß alle oder mehrere einfache Haupteffekte mit dem Maximum der einzelnen minimalen Umfänge für die einzelnen Tests auf einfache Haupteffekte getestet werden müssen.

Hager & Westermann (1983b, S.87ff) geben ein Beispiel für den CRF-22 Plan, in dem die WH eine spezielle ordinale Wechselwirkung impliziert. Diese wiederum ermöglicht die Ableitung von einseitigen Hypothesen über spezielle einfache Haupteffekte und Kontraste:

$$
\begin{aligned}
&H_1^1: \quad \mu_{11}-\mu_{12} > 0 \\
&H_1^2: \quad \mu_{21}-\mu_{22} > 0 \\
&H_1^3: \quad (\mu_{11}-\mu_{12})-(\mu_{21}-\mu_{22}) > 0 \quad .
\end{aligned}
\tag{7.43}
$$

Die letzte zugehörige Nullhypothese kann mit einem Permutationstest nur bei einem Modellwechsel zu einem 4-stufigen einfaktoriellen Plan getestet werden, denn mit den üblichen Reparametrisierungsbedingungen im zweifaktoriellen Zufallsplan ist H_1^3 äquivalent darstellbar als $4(\alpha\beta)_{11}>0$. Für einen Test auf Wechselwirkungen war aber kein Invarianz-Permutationstest möglich.

In Versuchsplänen mit wiederholten Messungen sind, wie bereits zuvor erwähnt, die insgesamt erwarteten intra- und interindividuellen Effekte getrennt zu spezifizieren (vgl. Bredenkamp, 1980, S. 92ff). Für den Split-Plot Versuchsplan gibt es damit ohne gezielte Hypothesen zwei globale Tests, einerseits für den intraindividuellen Split-Faktor (B) und dessen Wechselwirkungen mit dem Plot-Faktor (A) sowie für den interindividuellen Plot-Faktor. Für den ersten Globaltest ist bisher kein Invarianz-Permutationstest vorgestellt worden. Er läßt sich aber leicht konstruieren. In Analogie zu (3.59) muß die übliche Modellgleichung im Split-Plot Plan (3.61) verändert werden zu:

$$(7.44)\quad Y_{ih(i)j}=\mu+\beta_{ij}+\pi_{h(i)}+\varepsilon_{ih(i)j} \qquad \begin{array}{l} i=1,..,r \text{ Plotstufen} \\ j=1,..,c \text{ Splitstufen} \\ h(i)=1,..,n \text{ Replikationen} \end{array}$$

mit $\beta_{i.}=0$ für alle $i=1,..,r$.

Die Verteilungsfunktion des gesamten Fehlervektors muß lediglich invariant sein gegen Vertauschungen der Splitstufen und ganzer Blocke je Plotfaktorstufe $(Q_n^*=\times_{i=1}^{r}(S_c \wr S_n))$, denn die Nullhypothese

H: $\beta_{i1}=\ldots=\beta_{ic}$ für jedes $i=1,..,r$

gegen

A: $\beta_{ij}\neq\beta_{ij'}$ für ein Paar $j\neq j'$ $\in\{1,..,c\}$ und ein i

fordert ja Gleichheit der Split-Effekte getrennt für jede Plotstufe. Mit der Transformation

$$(KY)_{ih(i)j}=T_{ih(i)j}=Y_{ih(i)j}-\bar{Y}_{ih(i).} \quad \text{mit } K=I_{nr}\otimes(I_c-1/cJ_c)$$

eliminiert man die störenden Blockeffekte und erhält als zulassige Permutationen der transformierten Beobachtungen unter der Nullhypothese H: $Q_n=Q_n^*\cap S_c\wr S_{nr}=\times_{i=1}^{r}S_c\wr S_n$, also Vertauschungen innerhalb der Blöcke und ganzer Blöcke je Faktorstufe.

Als Teststatistik wählt man dann:

(7.45) $S(T) = \sum_{ij} n\bar{T}^2_{i.j}$ oder äquivalent $S(Y) = \sum_{ij} \bar{Y}^2_{i.j}$,

da die $Y_{ih(i).}$ invariant unter Permutationen aus Q_n sind.

Es handelt sich also formal um die Summe von r Teststatistiken für das verbundene c-Stichprobenproblem. Dann kann man in Analogie zu Bredenkamp (1980, S. 93f) vorgehen, wenn man wie in dessen Beispiel mir r=2, c=5 und einer Faktorstufe für den zweiten Plotfaktor 50% der Variabilität der abhängigen Variablen für die Blöcke veranschlagt und 10% zu Lasten des Splitfaktors zusammen mit der Wechselwirkung annimmt, d.h. $f^2_{B,AB}=.10/.40=.25$. Zur Bestimmung der minimal erforderlichen Anzahl an Blöcken benutzt man folgende Formel:

(7.46) $(n-1)r(c-1) = L_{B,AB}/f^2_{B,AB}$

$L_{B,AB}=22.74$ ist der Nichtzentralitätsparameterwert aus Tab. 9.4.2 in Cohen (1977) für $\alpha=\beta=.05$ und $u=r(c-1)=8$.

Damit ist $n=22.74\cdot4/8+1=12.37\cdot13$ und man benutzt insgesamt $rn=26$ Blöcke.

Der Permutationstest auf den Haupteffekt des Plotfaktors wurde bereits beschrieben (Willmes & Pyhel, 1981). Bekanntlich ist der Test auf Effekte des Plot-Faktors weniger teststark (vgl. Kirk, 1968, Kap. 8.11). Für denselben Effekt von 10% wie für B und AB zusammen, ergibt sich mit obiger Tafel $n=13\cdot4/2+1=27$, was insgesamt zu 54 erforderlichen Blöcken führen würde.

Bei eventuell auszuführenden Tests auf einfache Haupteffekte ist wieder zu bedenken, daß die Fehler-Quadratsummen des parametrischen F-Tests über alle Stufen des jeweils zweiten Faktors gebildet werden. Der Permutationstest bezieht sich aber stets nur auf die Beobachtungen bezüglich je einer Stufe des zweiten Faktors.

Abschließend sei noch einmal betont, daß alle diese Überlegungen zur Planung des im Experiment benötigten minimalen Stichprobenumfanges nur mehr oder weniger grobe Abschätzungen sein können, da nicht bekannt ist, wie stark die Abweichungen in der Größe des Nichtzentralitätsparameters bei Verletzung der ANOVA-Voraussetzungen ausfallen können.

(4) Die WH impliziert eine Alternative in einem multivariaten linearen Modell

Bei Verwendung des Spurkriteriums von Pillai aus (4.18) läßt sich für den parametrischen Fall leicht approximativ der Mindeststichprobenumfang bestimmen. Wie Bredenkamp (1980, S. 84ff) ausführlich für das unverbundene k-Stichprobenproblem dargestellt hat, gilt im d-dimensionalen Fall für den Gesamtstichprobenumfang bei Approximation der Verteilung des Spurkriteriums $V^{(s)}$ durch eine F-Verteilung (vgl. Kres, 1975):

(7.47) $N=L\cdot(s-V^{(s)})/(sV^{(s)}) + d-s+k$ mit $s=\min(k-1,d)$.

Der approximativ bestimmte Nichtzentralitatsparameter der F-Verteilung ist wiederum den Tabellen in Cohen (1977, Tab. 9.4.2) für vorgegebenes α,β und $f^2_d=V^{(s)}/(s-V^{(s)})$ zu entnehmen, jetzt aber mit $u=d(k-1)$ Zählerfreiheitsgraden der F-Statistik.

Wie im univariaten Fall ist die Formel (7.47) auf Tests linearer Hypothesen in anderen Versuchsplänen erweiterbar, wenn man allgemeiner definiert $s=\min(df_{Hypothese},d)$ und $k=df_{Hypothese}+1$ setzt. Der L-Wert ist bei vorgegebenem α, β und f^2_d für $u=df_{Hypothese}$ abzulesen (vgl. Bredenkamp, 1980, S. 88-90; Bredenkamp & Erdfelder, 1985). Nach Ausführung des

Experimentes kann in Verallgemeinerung des univariaten Falles f^2_d folgendermaßen geschätzt werden:

(7.48) $\hat{f}^2_d = v^{(s)}/(s-v^{(s)})$.

Andere Effektmaße sind bei Cramer & Nicewander (1979), Huberty (1972), Joe & Anderson (1969) sowie Moosbrugger & Steyer (1983), Sachdeva (1973), Smith & Stevens (1972) dargestellt.

Die Entscheidung bezüglich der WH trifft man dann wieder wie in (7.24) (vgl. Moosbrugger & Steyer, 1983, S. 199).

Falls spezifische zweiseitige Alternativen zur globalen Nullhypothese von der WH impliziert werden, sind diese für Permutationstests mit denselben Einschränkungen wie im univariaten Fall ausführbar, da auch hier zulässige Permutationen stets nur aus Vertauschungen bezüglich der von der speziellen Nullhypothese betroffenen Faktorstufen bestehen. Multiple Tests mit der T^2-Statistik nach Hotelling (vgl. (4.2)) mit reduziertem β-Niveau und je Vergleich spezifizierter Effektgröße sind leicht auszuführen. Einseitige Alternativen können mit Teststatistiken wie von Shuster & Boyett (vgl. Abschnitt 4.1) oder mit den multiplen sequentiellen Verfahren aus Abschnitt 5.3 überprüft werden.

7.4.3 Permutationstests für (streng) monotone Trendhypothesen bezüglich qualitativ abgestufter unabhängiger Variablen

Falls die verschiedenen Faktorstufen im (un)verbundenen k-Stichprobenproblem qualitativ abgestuft sind, ist keiner der in Abschnitt 7.4.2 beschriebenen Permutationstests für polynomiale Trends verwendbar. Aber auch bei quantitativer Veriation der Ausprägungen der unabhängigen Variablen ist es möglich, daß nicht der spezifische funktionale Zusammenhang von Interesse ist, sondern vielmehr nur eine Aussage über ordinale Beziehungen zwischen den Erwartungswerten (oder Medianen) der Beobachtungen für die verschiedenen Faktorstufen. In der Statistikliteratur sind für diesen Fall parametrische und nichtparametrische (Rang-)Tests gegen sog. geordnete Alternativen entwickelt worden (Barlow et al., 1972). Dabei handelt es sich häufig um Tests für Hypothesen der Form:

(7.49)

$$H: \mu_1=\mu_2=..=\mu_k$$

gegen die Alternative

$$A: \mu_1\leq\mu_2\leq..\leq\mu_k$$

mit mindestens einer echten Ungleichung

$\mu_i<\mu_{i+1}$ für ein $j\in\{1,..,k-1\}$ (Berenson, 1982b)

oder anders dargestellt mit der Bedingung

$\mu_1<\mu_k$ (Rothe, 1986).

Wie insbesondere Westermann & Hager (1983a,b) an verschiedenen Beispielen aus der psychologischen Forschung dargestellt haben, implizieren die dort behandelten WHen in der Regel nicht die obigen Alternativen als SHen sondern machen Aussagen uber spezifische Muster von Ungleichungen zwischen den Parametern, insbesondere von der Form einer strikten

Ordnung:

(7.50) $A: \mu_1 < \mu_2 < \ldots < \mu_k$.

Falls man nach dem signifikanten varianzanalytischen F-Test oder dem i.a. teststärkeren $\bar{E}_k^2$-Test von Bartholomew (s.u.) oder einem der nichtparametrischen Verfahren (vgl. Berenson, 1982b) eines der Verfahren zum multiplen zweiseitigen Vergleich von k-Faktorstufen verwendet, ist die Teststärke für die spezielle Alternative (7.50) in der Regel nicht besonders hoch, da auch viele andere Typen von Alternativen mit einer Ablehnung der Nullhypothese in (7.49) verträglich sind. Auch die in Abschnitt 5 vorgestellten multiplen Teststrategien haben denselben Nachteil. Westermann & Hager (1983a,b) und Hager & Westermann (1983, a,b,c) sowie Hager & Möller (1983) schlagen deshalb folgendes mehrschrittige Verfahren vor:

(i) Bei Festlegung einer i.a. relativ großen Effektstärke und kleinem β-Fehler wird zuerst ein 'Global'-Test ausgeführt. Für die Bestimmung des minimal erforderlichen Stichprobenumfanges im ANOVA-Fall sind die Überlegungen von Cohen (1977, Kap. 8.2.1) zur Spezifizierung einer (standardisierten) Spannweite zwischen den (geordneten) Erwartungswerten statt der globalen Angabe des Effektmaßes f^2 hilfreich. Entsprechende Tabellen sind für die $\bar{E}_k^2$-Teststatistik nicht publiziert. In vielen Fällen dürfte, falls die geordnete Alternative tatsächlich vorliegt, die Schätzung des Stichprobenumfanges für den üblichen F-Test deshalb eine Überschätzung sein. Wie Rothe (1986) aber gezeigt hat, gibt es Parameterkonstellationen, in denen der spezielle Test auf einen (linearen) Trend weniger teststark ist als der F-Test der ANOVA.

(ii) Bei signifikantem Ausgang des Tests aus (i) werden für den Fall einer strikten Ordnung k-1 einseitige Tests für die Hypothesen

(7.51) H^i: $\mu_i \geq \mu_{i+1}$ gegen A: $\mu_i < \mu_{i+1}$, i=1,..,k-1 ,

ausgeführt.

Als Tests eignen sich im parametrischen Fall einseitige t-Tests. Zur Schätzung des erforderlichen Stichprobenumfanges muß der β-Fehler je Test adjustiert werden. Die Effektstärke kann je Vergleich unter Umstanden unterschiedlich groß angesetzt werden. Falls dann unterschiedlich große Mindestumfänge resultieren, sollte i.a. ihr Maximum für alle Vergleiche gewahlt werden. Zu kontrollieren ist dann auch noch, ob dieser Stichprobenumfang für den Globaltest ebenfalls ausgereicht hätte. Ein Beispiel wird bei Hager & Möller (1983) behandelt.

Hager & Westermann (1983b, S. 79) betonen, daß es kein Verfahren gibt, um die Nullhypothese aus (7.51) mit einem einzigen Test direkt zu überprüfen. Solch eine Möglichkeit stellen aber sog. Rangierungsverfahren ('ranking procedures') als spezielle Mehrentscheidungsprobleme dar (Gibbons, Olkin & Sobel, 1977, insbesondere Kap. 12). Im einfachsten Fall mit Normalverteilungsannahme und bekannter Varianz für alle k Faktorstufen wird die erwartete Ordnung der Erwartungswerte spezifiziert. Sei das $\mu_{\pi(1)} \leq .. \leq \mu_{\pi(k)}$, wobei $\pi(1),..,\pi(k)$ eine Permutation der Indizes 1,..,k darstellt. Dann wird eine theoretisch erwartete Differenz $\delta_1^* = \mu_{\pi(i)} - \mu_{\pi(i-1)}$, l=1,..,k-1;i=k,..,2 spezifiziert. Im einfachsten Fall wird $\delta_1^* \equiv \delta$ angenommen.

Gesucht wird nun ein Verfahren, daß es gestattet, die Faktorstufen in eine Rangfolge zu bringen, für die die Wahrscheinlichkeit fur eine völlig richtige Rangfolge minde-

stens einen Wert von p* annimmt, falls jede der Differenzen δ_l mindestens den Wert δ^* hat:

(7.52) $P(\text{korrekte Rangfolge} \mid \delta_l \geq \delta^*,\ l=1,..,k-1) \geq p^*$.

Der erforderliche Stichprobenumfang ist z.B. aus Tabelle P.1 in Gibbons, Olkin & Sobel (1977) für eine spezifizierte Varianz zu entnehmen. Die gesuchte Prozedur besteht dann nur noch darin zu überprüfen, ob für die Rangordnung der Stichprobenmittelwerte gilt:

(7.53) $\bar{y}_{\pi'(1)} \leq .. \leq \bar{y}_{\pi'(k)}$ und $\pi'(i)=\pi(i)$ für alle $i=1,..,k$.

Die Rangordnung der Faktorstufenmittelwerte muß also identisch mit der angenommenen Ordnung sein.

Falls die Varianz aus den Daten gechätzt werden muß, gibt es zweischrittige Verfahren (vgl. Gibbons, Olkin & Sobel, 1977, Kap. 12.3), in denen zuerst mit einem kleineren Stichprobenumfang n für ein spezifiziertes δ^* und approximativ spezifiziertes σ^2 ein 'Schätzwert' $s^2=1/k\sum_i s_i$ bestimmt wird. In einem zweiten Schritt werden N-n Beobachtungen je Faktorstufe erhoben mit

(7.54) $N = \max(n, [2s^2h^2/\delta^{*2}])$.

Tabellen für eine approximative Wahl von n und verschiedene Werte von h in Abhängigkeit von n sind ebenfalls bei Gibbons et al (1977, Tab. P.2) angegeben. Die Prozedur selbst ist dann identisch mit der aus (7.53).

Diesen Verfahren ist gemeinsam, daß ein Fehler 1. Art nicht explizit festgesetzt ist. Wenn man jedoch ein theorietestendes Experiment bezüglich einer Ordnung von Mittelwerten überhaupt plant, ist m.E. die Wahrscheinlichkeit für einen Fehler 1. Art nicht besonders streng zu kontrollieren, da Gleichheit der Parameter kaum zu erwarten ist.

Im Folgenden sollen nun Permutationstests für Trendalternativen für das (un)verbundene k Stichprobenproblem vorgestellt werden.

Im verteilungsfreien, unverbundenen k Stichprobenproblem aus 3.1.5 (iii) mit $k \geq 3$ kann man wie im Normalverteilungsfall die $\bar{E}_k^2$-Statistik (Bartholomew, 1961; Barlow et al., 1972) verwenden (vgl. auch Berenson, 1982b; Berenson & Gross, 1981):

$$(7.55) \quad \bar{E}_k^2 = \sum_{i=1}^{k} n_i (Y_i^* - \bar{Y}_{..})^2 / \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 \quad .$$

Eine alternative Statistik wird von Pincus & Majumdar (1981) betrachtet. Da der Nenner unter der Nullhypothese des k-Stichprobenproblems permutationsinvariant ist, genügt es wieder im Permutationsmodell

$$(7.56) \quad \bar{E}^2{}_k = \sum_{i=1}^{k} n_i Y_i^{*2}$$

zu betrachten. Die Y_i^*, die sog. 'amalgamated means' (Roth, 1983), sind die unter der Alternative (7.49) zu bestimmenden Maximum-Likelihood Schätzer für die μ_i. Sie werden auch als 'isotone Regression' der $\bar{y}_{i.}$ bezeichnet. Zu den reellen Zahlen $x_1,\ldots,x_k$ heißt ein Vektor $(\theta_1^*,..,\theta_k^*)'$ die isotone Regression von $(x_1,..,x_k)'$, falls $\theta_1^* \leq \ldots \leq \theta_k^*$ ist und die Quadratsumme $\sum_i (x_i - \theta_i)^2$ unter der Nebenbedingung $\theta_1 \leq \ldots \leq \theta_k$ durch die θ_i^* minimiert wird.

Die Anzahl verschiedener Werte unter den θ_i^* wird oft mit $l(x_1,..,x_k)=l$ bezeichnet. Der bekannteste Algorithmus zur Ausführung der isotonen Regression ist der sog. 'pool adjacent violators'-Algorithmus (Barlow et al., 1972, S. 13ff und Beispiel 3.1 sowie Smith & Macdonald, 1983). In diesem Algorithmus werden die im Experiment enthaltenen Faktorstufenmittel $\bar{y}_{1.},\ldots,\bar{y}_{k.}$, die der Alternative (7.49)

widersprechen, so lange durch Zusammenfassen und Mitteln von benachbarten Werten in der Abfolge i=1,..,k verändert, bis die neuen $1 \le l \le k$ verschiedenen Werte y_i^* die Ordnungsbedingung erfüllen. Eine formale Darstellung der isotonen Regressionswerte ist mit

$$Y_i^* = \max_{1 \le l \le i} \min_{i \le l' \le k} \sum_{\nu=1}^{l'} n_\nu \bar{y}_{\nu\cdot} \Big/ \sum_{\nu=1}^{l'} n_\nu \quad , \; 1 \le l < l' \le k,$$

möglich (Chacko, 1963).

Je besser die beobachteten Faktorstufenmittel der Ordnungsbedingung genügen, umso größer wird (7.56) wie auch der Zähler in (7.55). Denn durch den evtl. notwendigen Mittelungsvorgang ('amalgamation process') wird die Spannweite der neuen y_i^* höchstens kleiner genauso wie ihre quadratischen Abweichungen vom festen Gesamtmittel der ursprünglichen Beobachtungen. Der exakte Permutationstest gegen die Alternative in (7.49) wird dann folgendermaßen ausgeführt: Für jede der $n_\cdot ! / \prod_{i=1}^{k} n_i !$ Permutationen von Werten zwischen verschiedenen Faktorstufen werden die y_i^* aus (7.56) mit dem oben erwähnten Algorithmus bestimmt und $\bar{E}_k^2$ berechnet. Die Nullhypothese des k- Stichprobenproblems wird bei vorgegebenem $0 < \alpha < 1$, zugunsten der obigen sog. einfachen Ordnungsalternative verworfen, falls der $p^{\geq}$-Wert $\le \alpha$ ist.

Williams (1971) hat zwei weitere Statistiken vorgeschlagen, die auch die Y_i^* benutzen. Diese lassen sich leicht unter Vernachlässigung von im Permutationsfall invarianten Größen auf diesen Permutationstestfall übertragen:

$$(7.57) \quad W_k(Y) = Y_k^* - \bar{Y}_{1\cdot} \qquad \text{bzw.} \qquad W_k^*(Y) = Y_k^* - Y_1^* \quad .$$

Eine andere Klasse von Teststatistiken haben Abelson & Tukey (1963) für gleiche Stichprobenumfänge vorgeschlagen. Ohne unter dem Permutationsmodell überflüssige Konstanten haben diese die Form (vgl. Barlow et al., 1972, Kap. 4.2):

$$(7.58)\quad T(Y) = \sum_{i=1}^{k} c_i \bar{Y}_{i.} \quad \text{mit} \quad \sum_{i=1}^{k} c_i = 0 \quad ,$$

wobei die Konstanten des linearen Kontrasts $c=(c_1,..,c_k)'$ so mit einem Algorithmus der quadratischen Programmierung zu bestimmen sind, daß das minimale Korrelationsquadrat von c mit $\mu=(\mu_1,..,\mu_k)'$, der hypothetischen Konfiguration der μ_i spezifiziert über lineare Nebenbedingungen, maximal wird. Bartholomew (1961) betrachtet auch den Fall ungleicher Stichprobenumfänge:

$$T^*(Y) = \sum_{i=1}^{k} c_i n_i \bar{Y}_{i.} \quad \text{mit} \quad \sum_{i=1}^{k} c_i n_i = 0 \quad .$$

Für den Fall einer einfachen Ordnung sind die c_i bestimmbar mit der Formel:

$$(7.59)\quad c_i = ((i-1)(1-(i-1)/k))^{1/2} - (i(1-i/k))^{1/2} \quad .$$

Weiterhin wurden für die Wahl der c_i folgende Vorschläge gemacht:

(7.60)
- (i) $c_1,..,c_k$ wie für einen orthogonalen linearen Kontrast $c_i=2i-k-1$
- (ii) $2c_1,c_2,..,c_{k-1},2c_k$ mit den c_i aus(i); 'linear-2'
- (iii) $4c_1,2c_2,..,2c_{k-1},4c_k$ mit den c_i aus (i); 'linear-2-4'.

Der Permutationstest wird mit der Teststatistik aus (7.58) wie im unverbundenen k-Stichprobenproblem ausgeführt. Levin et al. (1978) diskutierten die Verwendung der Statistik in (7.57), die einen nichtstandardisierten Korrelationskoeffizienten mit fixiertem c_i darstellt auch für weitere Klassen von Gewichten c_i für den Fall von k an einer Beobachtungseinheit gewonnenen Beobachtungen Y_i.

Die Teststarke dieser Testverfahren im Normalverteilungsfall gegen verschiedene Alternativen, die alle die Bedingung $\mu_1 \leq .. \leq \mu_k$ erfüllen wie:

(7.61)
$$\text{(i)}\quad \mu_{i+1}=\mu_i+c_i\ ,\ c_i>0,\ i=1,..,k-1$$
$$\text{(ii)}\quad \text{wie (i) mit } c_i=c$$
$$\text{(iii)}\quad \mu_1=...=\mu_{k-1}<\mu_k \text{ (`end gap`, `slippage`)}$$
$$\text{(iv)}\quad \mu_1<\mu_2=..=\mu_{k-1}<\mu_k\ ,$$

wurden von Bartholomew (1961), Berenson (1982b), Berenson & Gross (1981) und Marcus (1976) untersucht.

W_k und W_k^* haben für Alternativen der Form (iii) und (iv) größere Teststärke. Falls die c_i annahernd bekannt sind, hat die Abelson-Tukey Statistik gute Eigenschaften. Deren Teststärke nimmt jedoch stark ab, selbst wenn nur wenige der $c_i<0$ sind. Ansonsten ist $\bar{E}_k^2$ die bessere Wahl und auch teststärker als die ANOVA F-Statistik. Die Teststärke-Tabellen aufgrund von Monte-Carlo Untersuchungen von Berenson & Gross (1981) geben auch erste Hinweise zur Wahl eines geeigneten minimalen Stichprobenumfanges, um eine bestimmte Teststärke in dem Experiment zu erzielen.

Roth (1983) untersucht die Robustheit von $\bar{E}_k^2$ gegen Varianzheterogenität. Außerdem bezieht er in den Vergleich eine zum Test von Brown & Forsythe (1974) für das varianzheterogene k-Stichprobenproblem analoge Version von $\bar{E}_k^2$ ein. In diesem Fall gehen in die Bestimmung der Y_i^* nicht nur die ungleichen n_i sondern auch die n_i/s_i^2 ein, wobei die s_i^2 die Varianzen je Faktorstufe sind. Die robuste Version hat bei Verletzung der Varianzhomogenitatsbedingung eine großere Teststärke für mittlere und große n_i.

Als letztes soll eine Klasse von Testverfahren vorgestellt werden, in denen versucht wird, die übliche F-Statistik der ANOVA mit einer unter der globalen Nullhypothese von ihr unabhängigen Statistik zu kombinieren, die die Übereinstimmung aus hypothetisch angenommener einfacher oder partieller Ordnung und der Ordnung der im Experiment tatsächlich aufgetretenen Faktorstufenmittel erfaßt. Der erste Vorschlag stammt von Chassan (1960, 1962). Falls bei k Faktorstufen (und identischen Stichprobenumfängen) die apriori angenommene Ordnung $\mu_1 < .. < \mu_k$ sich in den Faktorstufenmitteln $\bar{y}_{i.}$ wiederfindet, ist der p-Wert nicht der zum berechneten F-Wert gehörige Wert p_F sondern:

(7.62) $p_{CH} = p_F/k!$,

der mit dem nominellen Niveau α verglichen werden muß. In Erweiterung dieser Idee für apriori Festlegungen von partiellen Ordnungen kann man p_F folgendermaßen verkleinern: Wenn man mit N_k die Anzahl von Anordnungen der μ_i bezeichnet, die mit der hypothetischen partiellen Ordnung verträglich sind, kann man mit

(7.63) $p_G = p_F . N_k/k!$

arbeiten. Beide Verfahren haben jedoch eine sehr geringe Teststärke, wenn nur eine oder wenige der vermuteten Ordnungsbeziehungen von den $\bar{y}_{i.}$ auch nur leicht verletzt werden oder wenn Gleichheit der Werte gilt (vgl. Barlow et al., 1972, S. 220f; Macdonald & Smith, 1983; Smith & Macdonald, 1983). Um diese Probleme zu vermeiden, haben Smith & Macdonald sowie unabhängig von ihnen Mantel (1983) folgendes Verfahren vorgeschlagen, das den generellen Ansatz der Kombination von unabhängigen Tests verwendet. Unter Gültigkeit der Nullhypothese des k-Stichprobenproblems sind die F-Statistik und eine Statistik, die die Übereinstimmung aus

vorhergesagter und tatsächlich aufgetretener Ordnung der μ_i erfaßt, voneinander unabhängig. Macdonald & Smith (1983) schlagen fur den Fall der einfachen Ordnung als diese Statistik den Kendall'schen Korrelationskoeffizienten τ zwischen den beiden Ordnungen vor, Mantel den Rangkorrelationskoeffizienten ρ von Spearman bzw. äquivalent die in umgekehrtem monotonem Zusammenhang stehende Summe der quadrierten Abweichungen der Rangplätze für beide Anordnungen (Mantel, 1983, S. 226). τ und ρ können für k Faktorstufen stets nur endlich viele Werte annehmen. Wenn man diese nach absteigender Größe $1=\tau_1>..>\tau_m=-1$ ordnet (für ρ analog), ist der p-Wert für den Test auf $\tau=0$ gegen $\tau>0$ bei beobachtetem τ_{m_o}, $m_m \leq m$:

$$p_\tau = \sum_{\nu=1}^{m_o} P(\tau=\tau_{m_\nu}) \quad .$$

Der stetig verteilte p-Wert des F-Tests p_F hat unter der Nullhypothese eine Gleichverteilung auf (0,1).

Zum Testen der Nullhypothese des k-Stichprobenproblems gegen

(7.64) A: $\mu_i \neq \mu_{i'}$, für mindestens ein Paar $i \neq i' \in \{1,..,k\}$

<u>oder</u> $\tau>0$

muß man die Verteilung der Zufallsvariablen

(7.65) $p_{F\cdot\tau} = p_\tau \cdot p_F$

bestimmen. Diese ist in Macdonald & Smith (1983, Formeln (8)-(10)) angegeben. Es läßt sich dann ausrechnen, wie groß bei einem festen Wert von τ der p-Wert p_F ausfallen darf, damit insgesamt $p_{F\cdot\tau}$ das nominelle α-Niveau des Tests für die Nullhypothese des k-Stichprobenproblems nicht uberschreitet. Mantel hat ein einfaches Rechenschema angegeben, in dem man das maximale p_F für einen Wert von τ bestimmen kann, damit von (7.65) das Niveau α eingehalten wird. Bei Macdonald & Smith (1983) gibt es dafür Tafeln für k'=2,..,9 von der

einfachen Ordnungsalternative betroffene Faktorstufen mit $k' \leq k$. Wenn $k' < k$ ist, können die Tabellen genauso wie für k benutzt werden; die F-Statistik hat jedoch weiterhin k-1 und N-k Freiheitsgrade.

Erste Teststärkevergleiche durch Macdonald & Smith von $\bar{E}_k^2$, dem Test von Chassan aus (7.62) und dem Test mit (7.65) für k=3,4 zeigen, daß $\bar{E}_k^2$ und der Test von Chassan die größte Teststärke besitzen, wenn die Ordnung perfekt erfüllt ist. Die Teststärke sinkt aber sehr schnell, wenn die vorhergesagte Ordnung deutlich verletzt ist. In diesem Fall ist besonders bei großen Effekten der zuletzt vorgestellte Test verglichen mit dem F-Test noch relativ teststark. Insbesondere ist die Eigenschaft des kombinierten Tests hervorzuheben, bei großen, stark von der Vorhersage abweichenden Effekten noch ein Verwerfen der Nullhypothese des k-Stichprobenproblems zu ermöglichen.

Dieser Ansatz einer Kombination von zwei unter der jeweiligen Nullhypothese unabhängigen Teststatistiken ist nicht auf den Fall eines Tests gegen eine einfache Ordnungsalternative beschränkt. Macdonald & Smith (1983) betrachten als weitere Beispiele die Vorhersage der Richtung von $k' \leq k-1$ unabhängigen Kontrasten von Bedingungen sowie die Vorhersage einer partiellen Ordnung. Im ersten Fall folgt die Anzahl N_c richtig vorhergesagter Richtungen der Kontraste unter der globalen Nullhypothese einer Binomialverteilung mit dem Parameter .5 und N=k'. Wenn man in den vorausgegangenen Formeln τ durch N_c ersetzt, verläuft alles weitere analog.

Im Fall einer vorhergesagten partiellen Ordnung werden Alternativen der Form

(7.66) $A: \{\mu_1, \mu_2, \ldots, \mu_s\} < \{\mu_{s+1}, \ldots, \mu_k\}$, $1 \leq s \leq k-1$,

betrachtet. Dann folgt unter der obigen Nullhypothese die Zufallsvariable

(7.67) X: Anzahl der Mittelwerte mit i=s+1,..,k größer als alle Mittelwerte mit i=1,..,s

einer hypergeometrischen Verteilung mit $k_1=|\{1,..,s\}|$, $k_2=|\{s+1,..,k\}|$; $k_1+k_2=k$. Wie für die einfache Ordnung sind für beide Fälle Tabellen des maximal zulässigen p_F für ein nominelles α-Niveau von .05 und .01, wenn N_c bzw. X einen bestimmten Wert n_c bzw. x annimmt, vorhanden. Weiterhin gibt es ein Computerprogramm von Macdonald (1984), mit dem man für jede beliebige andere Statistik, die ein Maß für eine Form von partieller Ordnung darstellt und die unter der Nullhypothese aus (7.64) unabhängig von der F-Statistik ist, den obigen kombinierten Test ausführen kann. Auf den Fall, daß als Alternative ein Anstieg der Parameterwerte im Mittel von Interesse ist, sei nur hingewiesen (vgl. Dykstra & Robertson, 1983; Robertson & Wright, 1982).

Der kombinierte Test wurde für den parametrischen Fall deshalb so ausführlich dargestellt, weil eine Übertragung dieses Konzeptes auf den Fall, in dem statt dem F-Test der entsprechende exakte oder Monte-Carlo Permutationstest verwendet wird, direkt möglich ist. Statt p_F ist $p^{\geq}$ (bzw. $p^{\leq}$) als p-Wert-Statistik zu verwenden. Allerdings ist der p-Wert im Permutationstest eine diskrete Zufallsvariable mit höchstens $|\mathcal{Q}_n|$ verschiedenen Vielfachen von $1/|\mathcal{Q}_n|$ als Realisation der $p^{\geq}$-Statistik, wobei $\mathcal{Q}_n$ die unter der Nullhypothese des Permutationstestproblems zulassige Permutationsgruppe bezeichnet. Für nicht zu kleine $|\mathcal{Q}_n|$ ist das Problem allerdings praktisch irrelevant. Einerseits sind Bindungen in den Teststatistikwerten dann wenig wahrscheinlich, sodaß die Verteilung von p einer diskreten Gleichver-

teilung auf $[1/|\Omega_n|, 1-1/|\Omega_n|]$ nahekommt. Um die Gleichverteilung zu erzwingen, kann man einen Vorschlag von Foutz (1980) aufgreifen und durch Multiplikation des Teststatistikwertes $S(y)$ aus (3.40) mit einer auf $[-\varepsilon,\varepsilon]$ gleichverteilten Zufallsvariablen mit hinreichend kleinem ε alle Werte der Teststatistik verschieden machen. Es ist auch möglich, statt des $p^{\geq}$-Wertes den in (2.7) definierten randomisierten p-Wert zu verwenden. Wenn man die Simulationsergebnisse von Macdonald & Smith als Anhaltspunkt nimmt, kann auch für den Permutationstest die Formel (7.30) zur approximativen Bestimmung des notwendigen Stichprobenumfanges verwendet werden. Denn für relativ große Effekte und nicht zu kleinen Stichprobenumfang ist die Teststärke des kombinierten Tests für relativ große Werte von τ nicht geringer als für den F-Test.

Auch das besonders von Hager & Westermann (1983a) sowie von Westermann & Hager (1983a,b) empfohlene Vorgehen, mit einseitigen t-Tests eine monotone Trendhypothese zu testen, ist direkt auf den Permutationstestfall übertragbar. Unter der wiederholt hervorgehobenen Einschränkung, daß die Planung des notwendigen Stichprobenumfanges für Permutationstests nur approximativ durch aufwendige Simulationsexperimente oder den Bezug auf den parametrischen Fall mit Normalverteilungsvoraussetzung möglich ist, können zum spezifizierten globalen Niveau α und β_i-Niveau für den i-ten Vergleich von 2 Faktorstufen die in Abschnitt 3.1.5 dargestellten Teststatistiken verwendet werden. Die tatsächlich im Experiment vorgefundene Effektstärke kann mit (7.25) geschätzt werden.

Fur den Fall des verbundenen k-Stichprobenproblems sind im parametrischen Fall ebenfalls mehrere Testverfahren gegen geordnete Alternativen vorgeschlagen worden, die sich alle auf folgende Form bringen lassen (vgl. Berenson, 1982a):

$$T(Y) = \sum_{j=1}^{k} c_j \bar{Y}_{.j}$$

(7.68) mit

(i) $c_j = 2j-k-1 \quad j=1,..,k$ (Kepner & Robinson,1984)

(ii) $c_j = 2-2k, c_2, ..., c_{k-1}, 2k-2$ (Linear-2 Kontraste)

(iii) $c_j = 4-4k, 6-2k, c_3, ..., 2k-6, 4k-4$ (Linear-4-2 Kontraste) .

Dieselbe Teststatistik kann für den Permutationstest verwendet werden. Nun sind lediglich die $(k!)^n$ Intrablockpermutationen der Beobachtungen zur Bestimmung von $p^{\geq}$ zu betrachten. Grobe Hinweise für die Planung des Stichprobenumfanges für den Test mit (7.68) kann man den Simulationsstudien von Berenson (1982a) zum Teststarkevergleich verschiedener parametrischer und nichtparametrischer Tests entnehmen. Im Permutationsmodell, in dem sich die Verteilung einer Teststatistik unter der Nullhypothese ja stets exakt oder approximativ mit einer Monte-Carlo Lösung bestimmen läßt, kann man leicht eine zu $\bar{E}_k^2$ aus (7.56) analoge Teststatistik konstruieren. Man bestimmt je Block i mit oder ohne vorhergehender Eliminierung der Blockeffekte die isotone Regression $\{Y_{ij}^*\}_i$ von $\{Y_{ij}\}_i$ und bildet:

$$(7.69) \qquad \bar{E}_{n,k}^2 = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} Y_{ij}^* \right)^2 .$$

Unter der Nullhypothese sind weiterhin alle $(k!)^n$ Intrablockpermutationen zulassig. Die Nullhypothese wird fur $p^{\geq} \leq \alpha$ zugunsten der Trendalternative verworfen.

Mit den Teststatistiken von Williams aus (7.57) sind in offensichtlicher Weise ebenfalls analoge Permutationstests konstruierbar.

Es ist auch möglich, nach Subtraktion der Blockmittel die isotone Regression $Y^*_{.j}$ der Faktorstufenmittel $\bar{Y}_{.j}$ zur Bildung einer Teststatistik zu nutzen:

$$(7.70) \qquad \bar{E}^2_k = \sum_{j=1}^{k} Y^{*2}_{.j} \quad .$$

Ansonsten wird wie zuvor beschrieben verfahren. Die Teststatistik in (7.70) sollte weniger anfällig gegen leichte Abweichungen von der einfachen Ordnung in einigen der n Blöcke sein.

Auch die kombinierten Tests von Macdonald & Smith (1983) sind übertragbar. Es können sogar dieselben Tabellen wie fur den unverbundenen Fall verwendet werden, denn unter der Nullhypothese sind weiterhin die F-Statistik oder die analoge Permutationsteststatistik und τ bzw. ρ oder N_c voneinander unabhängig.

Zum Abschluß sollen noch Verallgemeinerungen angedeutet werden. Marcus & Talpaz (1983) betrachten einen zweifaktoriellen Plan mit einer experimentell variierten und einer organismischen Variablen. Es soll getestet werden, ob in allen r Strata eine einfache Ordnung bezüglich der c Faktorstufen besteht. Ohne die Normalverteilungsannahme hat man als Modell:

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \qquad i=1,..,r;\ j=1,..,c;\ k=1,..,n$$

mit der Fehlerverteilung $F(x)=F(\pi x)$ für alle $\pi \in \times_{i=1}^{r} S_{cn,i}$, d.h. Vertauschbarkeit aller Variablen je Stratum unabhängig von Stratum zu Stratum. Getestet werden soll (vgl. Barlow et al., 1972, S. 215f):

(7.71) H: $\mu_{i1}=\ldots=\mu_{ic}$ für alle $i=1,..,r$

gegen

A: $\mu_{i1}\leq\ldots\leq\mu_{ic}$ und $\mu_{ij}<\mu_{ij'}$ für ein Paar $j\neq j'\in\{1,..,c\}$.

Als Teststatistik kann man dann die Summe von r Statistiken der Form (7.56) wählen:

(7.72) $$E^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} Y^{*2}_{i,ij} \quad .$$

Die $Y^{*2}_{i,ij}$ sind je Stratum i die isotone Regression der $\bar{Y}_{ij.}$. Die Permutationsverteilung von E^2 ist unter allen $((cn)!/n!c)^r$ Intrastratumpermutationen zwischen den c Faktorstufen zu bilden. Je zulässiger Permutation sind die $Y^*_{i,ij}$ aus (7.72) zu ermitteln und $p^{\geq}(y)$ mit der üblichen Testentscheidung zu bestimmen. Offensichtlich sind in analoger Weise auch die Teststatistiken aus (7.57) und (7.58) zu verallgemeinern.

Shorack (1967) und Kulatunga & Sasabuchi (1984) betrachten den zweifaktoriellen vollständigen Zufallsplan ohne Wechselwirkungsparameter mit einfacher oder mehrfacher gleicher Zellbesetzung unter dem Normalverteilungsmodell:

(7.73) $Y_{ijk} = \mu+\alpha_i+\beta_j+\varepsilon_{ijk}$, $i=1,..,r$; $j=1,..,c$; $k=1,..,n(\geq 1)$

Getestet werden soll die globale Nullhypothese:

(7.74) $H_{A,B}$: $\alpha_1=\ldots=\alpha_r$ und $\beta_1=\ldots=\beta_c$

gegen eine einfache Ordnungsalternative in beiden Faktoren simultan:

$A_{A,B}$: $\alpha_1\leq\ldots\leq\alpha_r$ und $\beta_1\leq\ldots\leq\beta_c$.

Als Teststatistik im Fall n=1 wird folgende 'Kombination' zweier $\bar{E}^2$-Statistiken vorgeschlagen mit $\{Y^*_{i.}\}$ bzw. $\{Y^*_{.j}\}$ den isotonen Regressionen von $\{\bar{Y}_{i.}\}$ bzw. $\{\bar{Y}_{.j}\}$:

(7.75) $\bar{E}^2_{A,B} = (c \sum_{i=1}^{r} (Y^*_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 + r \sum_{j=1}^{c} (Y^*_{\cdot j} - \bar{Y}_{\cdot\cdot})^2) / \sum_{i=1}^{r} \sum_{j=1}^{c} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$

Ein analoger Permutationstest läßt sich sofort angeben. Unter der Nullhypothese aus (7.72) sind, wenn man wie üblich für den gesamten Fehlervektor Vertauschbarkeit aller seiner rc Komponenten voraussetzt, alle y_{ij} vertauschbar. Für jede dieser (rc)! zulässigen Permutationen, die den Nenner aus (7.75) konstant lassen, müssen die isotonen Regressionen neu berechnet und der Zähler von $\bar{E}^2_{A,B}$ als Teststatistikwert ermittelt werden. $H_{A,B}$ wird verworfen, falls $p^{\geq}$ nicht größer als das nominelle Niveau α ausfällt. Die Teststärke dieses Tests ist bisher nur für den Spezialfall gleichabständiger Effekte der Faktorstufen mit der Teststärke des entsprechenden F-Tests verglichen worden. Ob bei kleinen Abweichungen von der einfachen Ordnung in einer oder in beiden Parametermengen die Überlegenheit des Tests mit (7.75) bestehen bleibt oder nur wenig unter die Teststärke des F-Tests sinkt, ist bisher nicht untersucht worden. Eine Verallgemeinerung auf den Fall mehrfacher gleicher Zellbesetzungen auch als Permutationstest ist in offensichtlicher Weise möglich. Eine Verallgemeinerung auf mehr als zwei gekreuzte Faktoren ist ebenfalls möglich, wie schon von Kulatunga & Sasabuchi (1984) angedeutet wird.

Von praktischem Interesse ist noch ein Trendtest für das Lateinische Quadrat. Wenn wie in Abschnitt 3.1.5 (iv) nur der Faktor C überprüft werden soll bzw. experimentell randomisiert wurde, kann man statt der Teststatistik (3.58), in die die Faktorstufenmittel der transformierten Beobachtungen eingehen, die Quadratsumme der isotonen Regressionswerte T^*_k dieser $T_k = 1/l \sum_{k(i,j)=K} T_{ij}$

(7.76) $\bar{E}^2 = \sum_{k=1}^{l} T^{*2}_k$

unter denselben bei Gültigkeit der Nullhypothese zulässigen Permutationen der $(1!)^2$ Zeilen- und Spaltenvertauschungen betrachten. Auch eine Teststatistik der Form

(7.77) $$T = \sum_{k=1}^{l} c_k T_k$$

mit Gewichten wie in (7.60) ist benutzbar ebenso wie die kombinierten Tests von Macdonald & Smith (1983).

Tests gegen Unterschiede in ordinalen Trends bezüglich eines Faktors für verschiedene Stufen eines zweiten Faktors sind ebenfalls vorgeschlagen worden (vgl. Huynh, 1981). Da es sich dabei um Tests handelt, die besonders sensitiv gegen bestimmte Formen von ordinalen Wechselwirkungen sind, werden sie erst im nächsten Abschnitt vorgestellt.
Alle Trendtests sind mit einer Maximumbildung über die univariaten Teststatistiken wie bei Shuster & Boyett (vgl. (4.6)) oder mit der multiplen sequentiellen Testprozedur aus Abschnitt 5.3 auf dem multivariaten Fall verallgemeinerbar.

7.4.4 Tests zur Analyse des Typs einer Wechselwirkung im linearen Modell

Besonders in Experimenten, die der Überprüfung einer WH dienen, ist die Analyse des Typs einer Wechselwirkung in der (den) von der WH implizierten SH(en) von Bedeutung. In einem (mindestens) zweifaktoriellen Experiment sind namlich haufig die aus der WH ableitbaren SHen nicht identisch mit dem Vorliegen irgendeiner Wechselwirkung, sondern es werden spezifische Vorhersagen über die Relationen von verschiedenen Wechselwirkungsparametern gemacht. Z.B. wird in der bereits zitierten Arbeit von Hager & Westermann (1983b, S. 89f) für einen zweifaktoriellen vollständigen Zufallsplan

eine spezielle ordinale Wechselwirkung vorhergesagt.

Wenn einer der (beiden) Faktoren in einem Experiment eine organismische oder experimentell nicht randomisierte unabhängige Variable ist, ist es wichtig festzustellen, ob eine (semi-)disordinale Wechselwirkung vorliegt, da ihr Vorhandensein die Populationsvalidität eines in einer WH formulierten Effektes der experimentell variierten unabhängigen Variablen einschränkt (Bredenkamp, 1980, S. 24ff, 1982, S. 59f; Hager & Westermann, 1983a, Kap. 4.3; Keppel, 1973, S. 316).

Eine generelle Lösung zur Feststellung des Typs einer Wechselwirkung mithilfe statistischer Tests ist auch für den ANOVA-Fall nicht vorhanden. Bredenkamp (1982) schlägt Verfahren vor, die sich auf einseitige polynomiale Kontraste beziehen. In der statistischen Methodenliteratur gibt es nur relativ wenige Publikationen, die sich mit einer solchen Fragestellung beschäftigen (Azzalini & Cox, 1984; Hirotsu, 1978, 1983; Smith & Macdonald, 1983). In einem anderen Vorschlag werden Teststatistiken, die für die Analyse monotoner Trends geeignet sind, miteinander verglichen (Huynh, 1981).

Im Folgenden wird nur auf Tests in zweifaktoriellen Versuchsplänen eingegangen. Die Überlegungen sind i.a. auf die Analyse von einfachen Wechselwirkungen zweier Faktoren auf einzelnen Stufen(kombinationen) eines (mehrerer) anderen(er) Faktors(en) erweiterbar. Eine Verallgemeinerung der Definitionen von ordinalen oder (semi-)disordinalen Wechselwirkungen auf mehr als zweifaktorielle Pläne ist nicht bekannt. Nachfolgend wird untersucht, inwieweit sich die Vorschläge für den ANOVA-Fall auf das Permutationsmodell übertragen

lassen.

Üblicherweise wird ein zweifaktorieller Versuchsplan ohne Faktoren mit wiederholten Messungen betrachtet, in dem entweder beide unabhängigen Variablen experimentell variiert sind oder eine unabhängige Variable eine (organismische) Kontrollvariable darstellt. Die in der Statistikliteratur vorgeschlagenen Verfahren testen das Vorhandensein bestimmter Typen von Wechselwirkungsparameterkonstellationen in einem eher explorativen Sinn (Hirotsu, 1983) mit für verschiedene Konstellationen unterschiedlich teststarken Verfahren.

In dem ANOVA-Modell

$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ mit unabhängigen, identisch normalverteilten Fehlern, i=1,..,r; j=1,..,c; k=1,..,n,

wurden als gerichtete Alternativen zur Nullhypothese der Additivität:

$$H: \mu_{i+1,j+1} - \mu_{i+1,j} - \mu_{i,j+1} + \mu_{ij} = 0 \quad , \; i=1,..,r-1; \; j=1,..,c-1$$

bzw.

$$H: \mu_{i'j'} - \mu_{i'j} - \mu_{ij'} + \mu_{ij} = 0 \quad i \neq i' \in \{1,..,r-1\}, \; j \neq j' \in \{1,..,c-1$$

folgende Varianten betrachtet (vgl. Hirotsu, 1978):

$$A`: \mu_{i+1,j+1} - \mu_{i+1,j} - \mu_{i,j+1} + \mu_{ij} \geq 0 \quad \text{für alle } i,j$$

und mindestens eine Ungleichung streng

oder äquivalent:

$$\mu_{i'j'} - \mu_{i'j} - \mu_{ij'} + \mu_{ij} \geq 0$$

und mindestens eine Ungleichung streng

(7.78) A``: (zweiseitige Version von A`)

$$\mu_{i'j'} - \mu_{i'j} - \mu_{ij'} + \mu_{ij} \geq 0 \qquad \text{für alle } i'>i;\; j'>j$$

oder

$$\mu_{i'j'} - \mu_{i'j} - \mu_{ij'} + \mu_{ij} \leq 0$$

und mindestens eine Ungleichung streng

A''': Für alle i i' 1,..,r gilt

$$\mu_{i,j+1}-\mu_{i',j+1}-\mu_{ij}+\mu_{i'j} \geq 0 \quad , \; j=1,..,c-1,$$

oder

$$\mu_{i,j+1}-\mu_{i',j+1}-\mu_{ij}+\mu_{i'j} \leq 0$$

und mindestens eine Ungleichung streng.

Im letzten Fall wird keine (natürliche) Ordnung bezüglich der Stufen des ersten (Gruppierungs-) Faktors angenommen. Die Alternative erfaßt Konstellationen, in denen mit wachsender Stufe j der zweiten (experimentellen) unabhängigen Variablen die Kontraste $\mu_{ij}-\mu_{i'j}$ einen wachsenden positiven oder negativen Wert annehmen. In der Alternative A' sind dagegen alle Fälle erfaßt, in denen mit wachsendem Index sowohl des ersten wir des zweiten Faktors eine größere Zunahme der Parameterwerte vorliegt.

Als eine 'schwächere' Version von A''' bzw. A'' werden von Hirotsu (1978) noch betrachtet:

$$A^{iv}: \; 1/c\sum_{j=1}^{m}(\mu_{ij}-\mu_{i'j})-(\bar{\mu}_{i.}-\bar{\mu}_{i'.}) \geq 0 \quad \text{für alle } i\neq i', \; m=1,..,c-1$$

oder

$$1/c\sum_{j=1}^{m}(\mu_{ij}-\mu_{i'j})-(\bar{\mu}_{i.}-\bar{\mu}_{i'.}) \leq 0$$

bzw.

$$A^{v}: \; 1/c\sum_{j=1}^{m}(\mu_{i+1,j}-\mu_{ij})-(\bar{\mu}_{i+1.}-\bar{\mu}_{i.}) \geq 0 \quad \text{für } i=1,..,r-1, \; m=1,..,c-1$$

oder

$$1/c\sum_{j=1}^{m}(\mu_{i+1,j}-\mu_{ij})-(\bar{\mu}_{i+1.}-\bar{\mu}_{i.}) \leq 0 \; .$$

Wie man sich anhand der graphischen Veranschaulichungen bei Hirotsu (1978) leicht überlegen kann, sind die mit den Alternativen A^{iv} und A^{v} verträglichen Muster von Parameterkonstellationen für Wechselwirkungen testende Fragestellungen zu unspezifisch. A'' und A''' und selbst A' sind bei ordinalen, semidisordinalen und disordinalen Wechselwirkun-

gen erfüllbar, was nachfolgende Konstellationen von Parameterwerten für den einfachsten Fall von r=c=2 Stufen je Faktor belegen:

	j=1	2	
2	1	3	2.0
1	0	1	0.5
	0.5	2	1.25

	j=1	2	
2	2	3	2.5
1	1	0	0.5
	1.5	1.5	1.5

	j=1	2	
2	0	2	1.0
1	1	0	0.5
	0.5	1	0.75

Die von Hirotsu (1978) vorgeschlagenen Tests von H gegen die Alternativen A`-A``` erlauben keine direkte Verallgemeinerung auf ein Permutationsmodell. Selbst bei Gultigkeit der Nullhypothese aus (7.78) sind keine gleichwahrscheinlichen Vertauschungen zwischen verschiedenen Stufen des Faktors B selbst innerhalb jeweils einer Stufe von A möglich. Wenn man wie in Abschnitt 3.1.5 die Haupteffekte durch Transformation eliminiert, bleiben unter dem Permutationsmodell für den Test auf Wechselwirkungen für die transformierten Zufallsvariablen T_{ijk} nur r!c! Vertauschungen ganzer Zeilen und Spalten neben den irrelevanten Permutationen innerhalb der Zellen übrig. Im Gegensatz zur Teststatistik (3.33) sind die von Hirotsu betrachteten Teststatistiken unter Vernachlassigung von irrelevanten Konstanten:

(7.79) $$S'_{AB}(T) = \sum_{i=1}^{r} \left(\sum_{j=1}^{c} c_j \bar{T}_{ij.} \right)^2$$

mit $c_j = j-(c+1)/2$ oder

$$c_j = (j(c-j))^{1/2} - ((j-1)(c-j+1))^{1/2}$$

(vgl. Abelson & Tukey,1963)

nicht mehr invariant unter den obigen zulässigen c! Vertauschungen der Stufen des zweiten Faktors. Für kleine Faktorstufenzahlen ist damit zwar wie für das Lateinische Quadrat (vgl. Abschnitt 3.1.5) prinzipiell ein Permutationstest moglich. Die erreichbaren Niveaus solch eines Tests gegen

$A^{\nu}-A^{\nu}$ beschränken sich jedoch auf Vielfache von 2/c!. Ohne die Quadratbildung in (7.79) ist auch ein Test gegen eine einseitige Alternative möglich. Statt des Kontrastes könnte auch $\sum_j T^*_{ij}$ mit der isotonen Regression der $\bar{T}_{ij.}$ je Stufe i des ersten Faktors betrachtet werden.

Im Modell (a) für den Split-Plot Plan (vgl. (3.61)) bleiben nach Eliminierung der Haupteffekte für eine zu (7.79) formal identische Teststatistik

(7.80) $S'_{AB}(T) = \sum_{i=1}^{r} (\sum_{j=1}^{c} c_j \bar{T}_{i.j})^2$ mit $T_{ih(i)j} = Y_{ih(i)j} - \bar{Y}_{ih(i).} - \bar{Y}_{..j} + \bar{Y}_{...}$

$n_.!c!/\prod_{i=1}^{r} n_i!$ zulässige nichttriviale Permutationen aus $S_{n_.} \wr S_c$ zum Testen eines Effektes der speziellen Wechselwirkungsparameter. Über die Vertauschungen ganzer transformierter Blöcke hinaus sind also im Unterschied zu Abschnitt 3.1.5 noch für alle $n_.$ Blöcke identische Vertauschungen von Split-Stufen hinzugekommen. Statt des linearen Kontrasten könnte erneut $\sum_j T^*_{i.j}$ verwendet werden. Auch hier wird die isotone Regression je Plotstufe i gebildet.

Während (7.79) prinzipiell symmetrisch bezüglich einer Vertauschung der beiden Faktoren ist, bleibt diese Eigenschaft im Split-Plot Plan nicht erhalten.

Für den zweifaktoriellen Blockplan mit Modell (3.39) findet man mit entsprechenden Überlegungen nach Eliminierung der Haupteffekte wie in Abschnitt 3.1.5 als Teststatistik:

(7.81) $S'_{AB}(T) = \sum_{i=1}^{r} (\sum_{j=1}^{c} c_j \bar{T}_{.ij})^2$ mit $T_{hij} = Y_{hij} - \bar{Y}_{hi.} - \bar{Y}_{h.j} + \bar{Y}_{h..}$

mit $(c!)^n$ nichttrivialen zulässigen Permutationen aus $\times_{i=1}^{n} (S_r \wr S_c)$. Anstelle der linearen Kontraste kann wieder die Summe der je Stufe des ersten Split-Faktors für jede zulás-

sige Permutation neu gebildeten isotonen Regression verwendet werden. Eine zu (7.81) 'symmetrische' Version mit Vertauschung von i und j ist ebenfalls verwendbar mit $(r!)^n$ zulässigen nichttrivialen Permutationen.

Huynh (1981) betrachtet für den Fall mit Normalverteilung eine spezielle Alternative zur Nullhypothese der Profilparallelität im Split-Plot Plan (vgl. (3.63)). Für den Fall zweier Plotstufen läßt sie sich darstellen als:

(7.82) $A: \delta_1 = \Delta_{11} - \Delta_{21} \leq \ldots \leq \delta_c = \Delta_{1c} - \Delta_{2c}$

mit

$$\Delta_{ij} = \mu_{ih(i)j} - \mu_{ih(i),j+1} \quad , \; j=1,..,c-1; \; h(i)=1,..,n_i$$

und mindestens einer strikten Ungleichung.

Die Veränderungen von Stufe zu Stufe des Splitfaktors sollen also nicht gleichförmig für beide Stufen des Plotfaktors sein, sondern sich mit steigender Splitstufe vergrößern.

Im parametrischen Fall werden die Differenzen $\bar{Z}_j = (\bar{Y}_{1.j} - \bar{Y}_{2.j})$ für $j=1,..,c$ betrachtet. Der Zähler der ANOVA-F-Statistik für den Test auf Wechselwirkung läßt sich für $r=2$ schreiben als

$$(n_1 n_2)/(n_1+n_2)) \cdot \sum_{j=1}^{c} (\bar{Z}_j - \bar{Z}_.)^2 \qquad \text{(vgl. Huynh, 1981).}$$

Ersetzt man die $\bar{Z}_j$ durch ihre isotonen Regressionen δ_j^*, erhält man den Zähler der in Analogie zum Test von Bartholomew (1961) gebildeten $\bar{E}^2$-Statistik (vgl. (7.55)).

Um einen entsprechenden Permutationstest konstruieren zu konnen, überlegt man, daß bei Gültigkeit der Nullhypothese der Profilparallelität aus (3.63) alle wie dort transformierten Blöcke untereinander vertauschbar sind. Dasselbe gilt für alle Stufen des Splitfaktors simultan über alle Blöcke hinweg nur dann, wenn S_c-Invarianz der Fehler je Block vorausgesetzt werden kann. Wenn die c Stufen des Splitfaktors nicht randomisiert zugewiesen wurden, ist diese

Voraussetzung in der Regel nicht haltbar (vgl. Abschnitt 3.1.5 (xi)).

Wenn kein monotoner Trend in den Differenzen aus (7.82) vorliegt, sollten die Differenzen der transformierten Beobachtungen nur wenig und zufällig um Null schwanken. In Analogie zu (7.55) kann man dann folgende Teststatistik wählen:

(7.83) $S(T) = \sum_{j=1}^{c} (\bar{T}_{1.j}-\bar{T}_{2.j})^{*2}$ $\qquad (\bar{T}_{1..}=\bar{T}_{2..}=0)$.

Die $(\bar{T}_{1.j}-\bar{T}_{2.j})^{*}$ bilden die isotone Regression der $(\bar{T}_{1.j}-\bar{T}_{2.j})$. Falls die Alternative gilt, wird es nur wenige der $(n_1+n_2)!/n_1!n_2!$ Permutationen der transformierten Blöcke geben, die zu einer größeren Quadratsumme der isotonen Regressionswerte führen. Die Nullhypothese (3.63) wird zugunsten von (7.82) verworfen, falls der $p^{\geq}$-Wert größer als das nominelle Niveau α ausfällt.

Bei Vertauschbarkeit der Blockfehler dürfen zusätzlich zu den Permutationen der transformierten Blöcke noch simultane Vertauschungen der Splitstufen auftreten, sodaß $(c!(n_1+n_2)!)/(n_1!n_2!)$ zulässige Permutationen zu betrachten sind. Alternativ zu (7.83) ist auch die Konstruktion einer dem Abelson & Tukey-Ansatz folgenden, zu (7.58) analogen Statistik möglich.

Die für den Fall von $r\geq 3$ Plotstufen von Huynh (1981) vorgeschlagenen Tests lassen sich ebenfalls direkt in Analogie zu (7.83) übertragen.

In der obigen Alternative A wird nur eine Hypothese über die Differenz von Parametern formuliert. Deshalb sind ordinale und bezüglich des Splitfaktors semi-disordinale Wechselwirkungen in gleicher Weise mit der Teststatistik

(7.83) zu entdecken. Bredenkamp (1982) hat nun für die ANOVA-Situation Verfahren vorgeschlagen, die auf die Bestimmung des Typs einer Wechselwirkung abzielen. Ihre Übertragbarkeit auf das Permutationsmodell soll abschließend untersucht werden. Um die prinzipiellen Überlegungen zu verdeutlichen, ist es ausreichend, den auch bei Bredenkamp vornehmlich behandelten Fall zu betrachten, in dem die Ordinalitat bzw. Disordinalität eines linearen Kontrastes für die Stufen des einen (Spalten-)Faktors bezüglich eines zweiten (Zeilen-)Faktors mit r Stufen in einem zweifaktoriellen Versuchsplan überprüft werden soll. Wie bei Bredenkamp werden die drei Versuchspläne CRF-rc, RBF-rc und SPF-r.c mit den in Abschnitt 3.1.5 angegebenen Modellgleichungen berucksichtigt. Es genügt sogar, nur den speziellen linearen Kontrast $c=(+1,-1)'$ für einen zweistufigen Spaltenfaktor zu behandeln. Weiterhin sei selbstverständlich vorausgesetzt, daß eine Situation vorliegt, in der die Prüfung der Ordinalität bzw. Disordinalität dieses Kontrastes bezüglich der zweiten unabhängigen Variablen überhaupt notwendig ist (vgl. Bredenkamp, 1982, S. 58-60). Schließlich sei noch angenommen, daß der Stichprobenumfang bei gleicher Zellbesetzung so gewahlt ist, daß die Gültigkeit einer Nullhypothese akzeptiert werden kann.

Zur uberprufung der Ordinalitat der Wechselwirkung hat man folgende Sequenz zu überprüfender Hypothesen:

(1) Vorhandensein des Kontrastes bei Mittelung über alle Stufen des zweiten (Zeilen-) Faktors:

$$H:^{(1)}: \sum_j c_j(\beta_j+\mu) \underset{(\geq)}{\leq} 0 \quad \text{gegen} \quad A^{(1)}: \sum_j c_j(\beta_j+\mu) \underset{(<)}{>} 0 \; .$$

Analog kann man auch den umgekehrten, in Klammern angegebenen Fall betrachten. Da bei Kontrasten $\sum_j c_j = 0$ ist, könnten die Hypothesen auch ohne den Parameter μ für das Gesamtmittel formuliert werden.

(7.84) (2) Falls $H^{(1)}$ für vorgegebenes α akzeptiert werden muß, erübrigt sich eine weitere Prüfung. Sei deshalb $A^{(1)}$ akzeptiert worden. Dann schlägt Bredenkamp vorab einen Globaltest auf Vorliegen einer Wechselwirkung für die Kontrastvariable mit dem zweiten Faktor vor, d.h. einen Test für:

$$H^{(2)}: \sum_j c_j \mu_{1j} = \sum_j c_j \mu_{2j} = \ldots = \sum_j c_j \mu_{rj} = \sum_j c_j (\beta_j + \mu)$$

oder äquivalent, da $\sum_j c_j = 0$:

$$\sum_j c_j (\alpha\beta)_{1j} = \ldots = \sum_j c_j (\alpha\beta)_{ij} = 0$$

gegen

$$A^{(2)}: \sum_j c_j (\alpha\beta)_{ij} \neq \sum_j c_j (\alpha\beta)_{i'j} \quad \text{für ein } i \neq i' \in \{1,..,r\} .$$

(3) Nur falls $H^{(2)}$ bei vorgegebenem α zu verwerfen ist, werden getrennt für jede Stufe des zweiten Faktors r einseitige Kontrastnullhypothesen getestet:

$$H_i^{(3)}: \sum_j c_j \mu_{ij} \underset{(\geq)}{\leq} 0 \quad \text{gegen} \quad A_i^{(3)}: \sum_j c_j \mu_{ij} \underset{(<)}{>} 0 \quad , \; i=1,..,r$$

oder äquivalent, da $\sum_j c_j = 0$:

$$\sum_j c_j (\beta_j + \alpha\beta_{ij}) \underset{(\geq)}{\leq} 0 \text{ gegen } \sum_j c_j (\beta_j + \alpha\beta_{ij}) \underset{(<)}{>} 0 \quad .$$

Falls alle zugehörigen Tests jeweils zum Niveau α aus (1) und (2) die jeweilige $H_i^{(3)}$ verwerfen, wird die Wechselwirkung als ordinal bezeichnet, ansonsten ist sie disordinal.

Zur Überprüfung der Disordinalität der Wechselwirkung sind folgende Schritte erforderlich:

(1)' Nur falls $H^{(1)}$ aus (7.84) bei vorgegebenem α zu akzeptieren ist, muß weiter getestet werden.

(7.85) (2)' Nur falls der Test der Hypothese $H^{(2)}$ aus (7.84) zur Annahme von $A^{(2)}$ führt, ist weiter zu testen.

(3)' Nur falls mindestens eine der Hypothesen $H_i^{(3)}$ aus (7.84) nicht verworfen werden kann, wird die Wechselwirkung als disordinal bezeichnet.

Es ist auch möglich, daß die WH eine bestimmte Struktur der disordinalen Wechselwirkung impliziert. Bredenkamp betrachtet den Fall, daß für bestimmte der $A_i^{(3)}$ das >-Zeichen und für die restlichen das <-Zeichen gelten soll. Dann kann mit einem einseitigen Test folgende Hypothese getestet werden:

$$H^{(4)}: \sum_i d_i \sum_j c_j \mu_{ij} \geq 0 \text{ gegen } A^{(4)}: \sum_i d_i \sum_j c_j \mu_{ij} < 0$$

(7.86)

$$\text{mit } d_i = \begin{cases} n_+ & \text{falls } A_i^{(3)}: \sum_j c_j \mu_{ij} < 0 \\ n_- & \text{falls } A_i^{(3)}: \sum_j c_j \mu_{ij} > 0 \end{cases}, \quad \sum_j d_i = 0$$

$$\text{und } n_+ = \{A_i^{(3)} \mid \sum_j c_j \mu_{ij} > 0\}; \quad n_- = \{A_i^{(3)} \mid \sum_j c_j \mu_{ij} < 0\} .$$

Wegen der letzten Beziehung für den Kontrast $d=(d_1,..,d_r)'$ hat die Nullhypothese $H^{(4)}$ folgende äquivalente Darstellung in den Wechselwirkungsparametern des linearen Modells:

$$H^{(4)}: \sum_i \sum_j d_i c_j (\alpha\beta)_{ij} \geq 0$$

(7.87) gegen

$$A^{(4)}: \sum_i \sum_j d_i c_j (\alpha\beta)_{ij} < 0 .$$

Im Normalverteilungsfall lassen sich, wie von Bredenkamp (1982) ausgeführt, für alle drei betrachteten zweifaktoriellen Pläne leicht t- bzw. F-verteilte Teststatistiken angeben. Analoge Tests sind möglich, wenn man die Rolle von Zeilen- und Spaltenfaktor vertauscht. Smith & Macdonald (1983) haben für $H^{(4)}$ ebenfalls Vorschläge gemacht, in denen der F-Test für die Wechselwirkungsparameter durch die Spezi-

fizierung von Vorzeichen von Kontrasten in den Wechselwirkungsparametern verschärft wird.

Unter den entsprechenden Permutationsmodellen ergeben sich dagegen Komplikationen. Es dürfen ja nur diejenigen transformierten Beobachtungen permutiert werden, deren gemeinsame Verteilung bei Gültigkeit der jeweiligen Nullhypothesen $H^{(1)} - H^{(3)}$ bzw. $H^{(4)}$ unter diesen Vertauschungen permutationsinvariant ist.

Im CRF-rc gibt es keine Möglichkeit, im Modell mit Wechselwirkungsparametern eine Transformation zu finden, die bis auf die β_j alle Effektparameter eliminiert. Es ist nur möglich, mit einem Test die Hypothesen

(7.88) $H_i^{(1)*}: \sum_j c_j(\beta_j+(\alpha\beta)_{ij}) \leq 0$ gegen $A_i^{(1)*}: \sum_j c_j(\beta_j+(\alpha\beta)_{ij}) > 0$ für alle $i=1,..,r$

zu testen. Für c=2 hat man als Teststatistik:

(7.89) $S(Y) = \sum_{i=1}^{r} \bar{Y}_{i1.}$ oder allgemein: $S(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} c_j \bar{Y}_{ij.}$.

Unter $\bigcap_i H_i^{(1)*}$ dürfen unabhängig von Stufe zu Stufe des 'Zeilen'-Faktors alle Beobachtungen vertauscht werden $(\pi \in \times_{i=1}^{r} S_{cn})$. Die Nullhypothese wird verworfen, falls der $p^{\geq}$-Wert $\leq \alpha$ ist. In diesem Fall hat man keine Probleme, wie im parametrischen Fall fortfahren zu dürfen. In Schritt (2) oder (3) in (7.84) müßte sich ja zeigen, ob die Ablehnung vornehmlich durch die Beiträge der β_j oder der $(\alpha\beta)_{ij}$ zustande gekommen ist. Wenn aber $p^{\geq} > \alpha$ ausfällt, ist nicht entscheidbar, ob evtl. eine disordinale Wechselwirkung für den Testausgang verantwortlich war. Wie man aus (7.84) entnehmen kann, läßt sich die Hypothese einer 'Wechselwirkung des Kontrastes mit dem zweiten Faktor' ausschließlich unter Verwendung von Wechselwirkungsparametern ausdrücken. Um sie mit einem Invarianz-Permutationstest zu testen, müßten also die Haupteffekte durch eine Transformation eliminiert

werden. Diese würden aber auch die Wechselwirkungsparameter eliminieren. Lediglich Schritt (3) in (7.84) wäre problemlos mit r einseitigen Permutationstests anhand der Teststatistiken

$$S_i(Y) = \sum_{j=1}^{c} c_j \bar{Y}_{ij.}$$

und Ausführung aller (cn)!/cn! nichttrivialen Permutationen innerhalb jeder Stufe i des zweiten Faktors durchführbar. Jeder der r $p^{\geq}_i$-Werte wird mit α verglichen. Während im parametrischen Fall in den Nennern der t-verteilten Teststatistiken Fehlerschätzungen gebildet über alle Stufen von i verwendet werden, sind in Permutationstests stets nur Vertauschungen über Beobachtungen je einer Plotstufe unter der jeweiligen $H_i^{(3)}$ zulässig.
Auch die Hypothese über die Struktur einer disordinalen Wechselwirkung in (7.87) wäre nur bei einem Modellwechsel zu einem k=rc Stichprobenproblem testbar.

Im SPF-r.c ist ebenfalls nur eine Hypothese der Form (7.88) im ersten Schritt aus (7.84) testbar. Man betrachtet die um die Blockmittel transformierten Beobachtungen $T^*_{ih(i)j} = Y_{ih(i)j} - \bar{Y}_{ih(i).}$ und wählt als Teststatistik:

(7.90)
$$S(T^*) = \sum_{i=1}^{r} \sum_{j=1}^{c} c_j \bar{T}^*_{i.j} \quad .$$

Unter $\bigcap_i H_i^{(1)*}$ dürfen unabhängig von Plotstufe zu Plotstufe die transformierten Beobachtungen je Block über die Splitstufen vertauscht werden ($\pi \in \times_{i=1}^{r} (S_c \wr S_n)$) , sodaß $((c!)^n)^r$ zulässige Permutationen vorhanden sind. Es ergeben sich dieselben Interpretationsprobleme wie zuvor, falls $p^{\geq} > \alpha$ ausfällt. Der globale Test auf Wechselwirkung des Kontrastes wird mit den wie in Abschnitt 3.1.5 transformierten Beobachtungen $T_{ih(i)j} = Y_{ih(i)j} - \bar{Y}_{ih(i).} - \bar{Y}_{..j} + \bar{Y}_{...}$ ausgeführt. Eine

geeignete Teststatistik ist:

$$(7.91) \qquad S(T) = \sum_{i=1}^{r} \left(\sum_{j=1}^{c} c_j \bar{T}_{i.j} \right)^2 \qquad \left(\bar{T}_{..j} = 0 \quad \text{für } j=1,..,c \right) .$$

Falls der $p^{\geq}$-Wert, berechnet unter allen $\pi \in S_{rn} \otimes S_c$, d.h. Vertauschungen ganzer transformierter Blocke und Vertauschungen der Splitstufen simultan über alle rn Blöcke, nicht größer als α ausfällt, müssen im dritten Schritt getrennt für jede Plotstufe einseitige Kontraste für die $T^*_{ih(i)j}$ mit folgenden Statistiken überprüft werden:

$$(7.92) \qquad S_i(T^*) = \sum_{j=1}^{c} c_j \bar{T}^*_{i.j} .$$

Wie im einfaktoriellen Blockplan werden unabhängig von Block zu Block unter der Plotstufe i alle c! Vertauschungen der Splitfaktorstufen ausgeführt. Eine ordinale Wechselwirkung wird nur dann als gegeben angesehen, wenn alle einseitigen $p^{\geq}$-Werte jeweils nicht größer als α ausfallen.

Eine Strukturhypothese der Form (7.86) ist wegen der Existenz eines Tests für die Wechselwirkungshypothese in (7.84) direkt angebbar. Man betrachtet alle unter der globalen Nullhypothese (3.62) zulässigen Permutationen $\pi \in S_{rn} \otimes S_c$ und bestimmt $p^{\geq}$ für die mit den transformierten Beobachtungen $T_{ih(i)j}$ gebildete Teststatistik:

$$(7.93) \qquad S(T) = \sum_{i=1}^{r} d_i \sum_{j=1}^{c} c_j \bar{T}_{i.j} .$$

Auch im RBF-rc ist die erste Hypothese aus (7.84) nur unvollkommen mit einem Permutationstest uberprüfbar. Man betrachtet die je Block transformierten Beobachtungen $T^*_{hij} = Y_{hij} - \bar{Y}_{hi.}$ und wählt in Analogie zum vorausgegangenen Fall als Teststatistik:

$$(7.94) \qquad S(T^*) = \sum_{i=1}^{r} \sum_{j=1}^{c} c_j \bar{T}^*_{.ij} .$$

Unter der Nullhypothese (7.88) dürfen unabhängig von Block zu Block die transformierten Beobachtungen über die Stufen des interessierenden Faktors separat je Stufe des zweiten Faktors vertauscht werden ($\pi \in \times_{i=1}^{r}(S_c \wr S_n)$) . Man hat wie im Split-Plot Plan $(c!)^{n})^{r}$ zulässige Permutationen. Wiederum ergeben sich für $p^{\geq} > \alpha$ Interpretationsprobleme.

Der Globaltest auf Wechselwirkungen für den Kontrast mit dem zweiten Splitfaktor erfordert transformierte Beobachtungen $T_{hij} = Y_{hij} - \bar{Y}_{hi.} - \bar{Y}_{h.j} + \bar{Y}_{h..}$, mit denen folgende Teststatistik zu bilden ist:

(7.95) $S_i(T) = \sum_{i=1}^{r} (\sum_{j=1}^{c} c_j \bar{T}_{.ij})^2$ (da $\bar{T}_{..j} = 0$ für alle $j=1,..,c$) .

Mit allen auch für den üblichen Test auf Wechselwirkungen im zweifaktoriellen Blockplan zulässigen Permutationen $\pi \in S_n \wr (S_r \otimes S_c)$ wird, falls $p^{\geq} \leq \alpha$ ist, zum nächsten Schritt, der Analyse der einfachen Kontrasteffekte je Splitstufe i, übergegangen. Dort testet man mit den T^*_{hij} mit folgenden Teststatistiken:

(7.96) $S(T^*) = \sum_{j=1}^{c} c_j \bar{T}^*_{.ij}$.

Die zulässigen Permutationen sind dieselben wie im einfaktoriellen Blockplan $(\pi \in S_c \wr S_n)$.

Eine ordinale Wechselwirkung wird nur dann angenommen, falls alle r einseitigen $p^{\geq}$-Werte nicht größer als das vorgegebene α-Niveau sind.

Wie im Split-Plot Plan ist auch hier eine Strukturhypothese des Typs (7.86) testbar. Statt des Globaltests in (7.95) ist nun unter denselben dort zulässigen Permutationen die folgende Teststatistik geeignet:

(7.97) $S(T) = \sum_{i=1}^{r} d_i \sum_{j=1}^{c} c_j \bar{T}_{.ij}$.

Wie zuvor in Abschnitt 3 bereitet nur der zweifaktorielle vollständige Zufallsplan prinzipielle Schwierigkeiten bei der Prüfung der Hypothesen aus (7.84) bzw. (7.85). Inwiefern die Kombination von r Statistiken zur Überprüfung der Hypothese $H^{(1)}$ zu einer geringeren Teststärke führt, ist nicht untersucht. Dieser Ausweg stellt aber im Permutationsmodell die einzige Möglichkeit dar, einen globalen Effekt des interessierenden Kontrastes zu testen.

Der Einsatz von Permutationstests ist nicht auf den Fall eines einfachen Kontrastes beschränkt. Auch die Kombination von mehreren solcher Tests wie in dem fiktiven Beispiel von Bredenkamp (1982, S. 72ff) zur Prüfung einer umgekehrt U-förmigen Lernleistungshöhe in Abhängigkeit vom Motivationsniveau ist prinzipiell möglich, wenn für den Faktor 'Motivationsniveau' wiederholte Messungen vorliegen.
Abschließend sei noch auf die 'quasi-statistischen' Vorschläge zur Prüfung des Typs einer Wechselwirkung von Hager (1982), Hager & Westermann (1983a, b, S. 87ff) und Henning & Muthig (1979) hingewiesen.

8. PRAKTISCHE AUSFÜHRUNG VON RANDOMISIERUNGS- UND PERMUTATIONSTESTS

8.1 Überblick über bekannte Verfahren

In der kritischen Würdigung durch Bradley (1968, Kap.4.3) werden Randomisierungstests trotz ihrer sehr guten mathematischen Eigenschaften als selten praktikabel und oft nicht durchführbar bezeichnet. Bredenkamp (1980, S. 20) spricht von einem 'unerhörten Rechenaufwand'. Fisher (1935) hat Permutationstests bereits als unpraktisch abgelehnt. Ohne einen Computer waren Fisher, Pitman, Welch u.a. tatsächlich nur für kleine (Gesamt-) Stichprobenumfänge in Ein-, Zwei- oder k-Stichprobenproblemen in der Lage, den exakten p-Wert in einem Permutationstest zu bestimmen. Auch Überlegungen zur geschickten Bestimmung der zulässigen Permutationen, die die 'extremsten' Teststatistikwerte ergeben (vgl. z.B. Green, 1977), führten nicht wesentlich weiter, denn die benötigte Rechenzeit für eine vollständige Berechnung aller zu zulässigen Permutationen gehörenden Werte der Teststatistik steigt sehr schnell mit wachsendem Gesamtstichprobenumfang (vgl. z.B. Pagano & Tritchler, 1983; Streitberg, 1981). Folgende Auswege aus diesem Dilemma wurden versucht:

(i) Wie zuvor bereits dargestellt (vgl. z.B. Abschnitt 6.4), konnte für verschiedene Testprobleme gezeigt werden, daß die ersten zwei bis vier exakten Momente der Permutationsverteilung der jeweiligen Teststatistik oder einer geeigneten (monotonen) Transformation von ihr für wachsenden Stichprobenumfang gut mit den entsprechenden Momenten der parametrischen Teststatistiken teilweise nach Adjustierung der Freiheitsgrade übereinstimmen (vgl. z.B.

Pitman, 1937a, 1938; Scheffe, 1959, Kap. 9.3). Unter diesem Blickwinkel ließ sich ein ANOVA-F-Test als Approximation an den entsprechenden Permutationstest auffassen (Bredenkamp, 1980).

Konzeptuell identisch ist die Approximation der (finiten) Überschreitungswahrscheinlichkeit für MRPP-Statistiken anhand der ersten drei oder vier Momente der exakten Permutationsverteilung der Teststatistik, wie sie in Abschnitt 4.7 beschrieben wurde.

(ii) Für das Ein-, Zwei- und k-Stichprobenproblem wurden Approximationen der exakten Permutationsverteilung mit Hilfe von sog. Reihenentwicklungen (Edgeworth-Reihen bzw. Sattelpunkt-Approximationen) vorgeschlagen (John & Robinson, 1983b; Robinson, 1978, 1980, 1982). Insbesondere werden Formeln für die approximative Bestimmung des p-Wertes hergeleitet. Je nach angestrebter Genauigkeit der Näherung wird der Wert der Verteilungsfunktion der Standardnormalverteilung an der Stelle des - gegenüber den Formeln in Abschnitt 3.1.5 geeignet standardisierten - Teststatistikwerten korrigiert. Diese Korrekturausdrücke berücksichtigen i.a. die 3. und 4. empirischen Momente der Beobachtungen. Die von den Autoren untersuchten Beispiele zeigen eine gute Übereinstimmung von exakten und approximativen p-Werten schon für recht kleine Stichprobenumfänge.

Anhand dieser Approximationen sind auch approximative ein- wie zweiseitige Konfidenzintervalle für das Ein- und Zweistichprobenproblem bestimmbar (John & Robinson, 1983a).

Mit denselben mathematischen Techniken sind auch Abschätzungen für die exakte Teststärke gegen Lagealternativen zu gewinnen. Explizit sind bisher nur Formeln für das unverbun-

dene Zweistichprobenproblem von John & Robinson (1983b) publiziert worden. Die Autoren deuten an, daß für das Ein- und k-Stichprobenproblem analoge Verfahren verwendbar sein müßten. In den eigenen Computer-Programmen sind die approximativen p-Werte und Teststärkeberechnungen als Option enthalten (vgl. Abschnitt 8.4.1). Die rechnerisch noch aufwendigeren sog. Sattelpunkt-Approximationen (Robinson, 1982) sind dort nicht berücksichtigt.

(iii) Für das Ein-, Zwei- und k-Stichprobenproblem wurden Möglichkeiten gefunden, die gesamte exakte Permutationsverteilung der jeweiligen Teststatistik bzw. den zugehörigen p-Wert zu bestimmen, ohne daß alle (zulässigen) Permutationen der ursprünglichen Beobachtungen generiert und für jede von ihnen der Wert der jeweiligen Teststatistik neu berechnet werden müßten. Der erste im englischsprachigen Bereich publizierte (kombinatorische) Algorithmus für das (un-)verbundene Zweistichprobenproblem stammt von Green (1977); siehe aber auch schon Heiler & Weichselberger (1969). Durch den Vergleich von 'Partialsummen' aus den Beobachtungen mit dem Teststatistikwert (3.38) bzw. (3.39) und eine geschickte Indizierung kann ein sehr großer Anteil (90% und mehr) aller möglichen Summenbildungen entfallen. Für Probleme mit weniger als ca. 3 Millionen zulässigen Permutationen wird ermittelt, ob der p-Wert kleiner als ein vorgegebenes α-Niveau ausfällt. Die (CPU-) Rechenzeiten sind dann nie länger als 20 Sekunden (Programmiersprache FORTRAN, DEC10-Computer).

Einen anderen (kombinatorischen) Ansatz wählen John & Robinson (1983a) für dieselben beiden Testprobleme. Kernstück des Algorithmus ist wiederum eine Überlegung, die Menge aller zu berechnenden Summen in (3.38) bzw. (3.39)

moglichst klein zu halten. Die Idee hat Ähnlichkeit mit dem Verfahren der sog. typischen Teilstichprobenwerte (typical subsample values) von Hartigan (1969; siehe auch Forsythe & Hartigan, 1970). Für Gesamtstichprobenumfänge von weniger als 25 sind die Rechenzeiten nie größer als 5 Sekunden (FORTRAN CDC CYBER 76). Es ist auch leicht möglich, ein Konfidenzintervall fur den Unterschied in den Lageparametern der beiden Verteilungen bzw. im verbundenen Zwei-Stichprobenproblem für den Lageparameter zu bestimmen.

Von größerer Allgemeinheit in den Anwendungsmöglichkeiten sind die Algorithmen von Pagano & Tritchler (1983) und Streitberg & Röhmel (1983, 1986). Beiden liegt ein sog. Verschiebe-Algorithmus (Shift-Algorithmus, siehe auch Oldenbürger, 1981) zugrunde, der mit wachsendem n nur zu einem polynomialen Anstieg in der Anzahl benötigter Rechenoperationen führt. Mit beiden Algorithmen wird die gesamte Verteilung der Teststatistik im Ein- und Zweistichprobenproblem (auch mit Stratifizierung) bestimmt. Für den k-Stichprobenfall gibt es bei Pagano & Tritchler Ansatze zu einer allgemeinen Lösung, bei Neumann (1985) einen Vorschlag für das verbundene k-Stichprobenproblem. Während Pagano & Tritchler (1983) den 'Umweg' über die sog. charakteristische Funktion der diskreten Teststatistikzufallsvariablen mit FORTRAN-Routinen wählen, konnen Streitberg & Röhmel (1983, 1986) die Verschiebe-Operationen in der Programmiersprache APL direkt ausführen (vgl. Abschnitt 8.3). Ein vergleichbarer FORTRAN-Algorithmus für das (un-)verbundene Zweistichprobenproblem stammt von Zimmermann (1985a, b). Die Umwandlung erfordert ebenfalls nicht viele Programmzeilen (vgl. Anhang).

Nachteilig bei diesen Algorithmen ist, daß nur (positive) ganzzahlige Werte verwendet werden durfen und daß es bei größeren Gesamtstichprobenumfangen auch bei Computern mit dynamischer Speicherplatzzuweisung zu Speicherplatzproblemen kommen kann.

Für einen großen Teil der in den Abschnitten 3 - 7 entwickelten uni- und multivariaten Permutationstests für komplexere Versuchspläne sind solche Algorithmen nicht bekannt.

(iv) Da schon die Verallgemeinerung des Zweistichprobenfalls auf den k-Stichprobenfall bei Pagano & Tritchler (1983, S. 439f) für beliebiges k > 3 sehr aufwendig erscheint, die benötigte Rechenzeit sehr schnell in k wächst und sich die Teststatistiken i.a. nicht als lineare Größen darstellen lassen, muß man besonders für multivariate Testprobleme doch Lösungswege über die explizite Betrachtung von (zulässigen) Permutationen der Komponenten des gesamten Beobachtungsvektors anstreben.

Dabei sind 'schnelle' Algorithmen zur sukzessiven systematischen Generierung aller zulässigen Permutationen von besonderer Bedeutung. Einen Überblick und einen schnellen FORTRAN-Algorithmus für das k-Stichprobenproblem findet man bei Berry (1982). Pyhel (ohne Jahr) hat ebenfalls geeignete FORTRAN-Algorithmen, die neue Permutationen möglichst lediglich durch Inversionen erzeugen, entwickelt, die für alle in den Abschnitten 3 - 7 vorgeschlagenen Permutationstestverfahren geeignet sind (vgl. Abschnitt 8.3). Prinzipiell andern diese Algorithmen aber nichts an dem sehr schnellen Wachstum der erforderlichen Teststatistikberechnungen mit steigendem Gesamtstichprobenumfang.

(v) Dwass (1957) und unabhängig von ihm Barnard (1963) in einem Diskussionsbeitrag haben vorgeschlagen, lediglich eine (große) Zufallsstichprobe aus der Menge aller zulässigen Permutationen zu benutzen, um einen vertretbaren Rechenaufwand für beliebige Gesamtstichprobenumfänge zu gewährleisten (vgl. auch Gabriel & Hall, 1983; Shuster, 1978). Während Dwass mehr den Fall diskreter (bedingter) Prüfgrößenverteilungen betrachtet, behandelten Barnard & Hope (1968) den Fall einer stetigen Prüfgrößenverteilung. Verschiedentlich wird dieses Vorgehen als Monte-Carlo-Methode (Hope, 1968; Pyhel, 1980a, b; Rubinstein, 1981) bezeichnet. Dieser Name soll andeuten, daß mit Hilfe eines Simulationsexperimentes (stochastisch unabhängige) Realisationen der jeweiligen Teststatistik unter Gültigkeit der Nullhypothese gewonnen werden. Ohne diese Bezeichnung hatten Eden & Yates (1933) dieses Prinzip zur Überprüfung der Robustheit einer Teststatistik gegen Verletzungen der Normalverteiltheit schon verwendet.

Die Charakterisierung als approximativer Permutationstest (Edgington, 1969b) ist nicht korrekt (vgl. Pagano & Tritchler, 1983), da der Test weiterhin ein exakter Test zum Niveau α bleibt (vgl. z.B. Hope, 1968). Die notwendige Einbuße an Teststärke ist dagegen für nicht zu kleine Zufallsstichproben von 10.000 und mehr nur sehr gering. Weiterhin kann gezeigt werden, daß die Teststärke des Monte-Carlo Permutationstests mit wachsendem Umfang der Zufallsstichprobe nicht abnehmen kann (vgl. Hope, 1968, Jöckel, 1982, 1986a, b).

Baker & Collier (1966a) haben ein Monte-Carlo Permutationstestprogramm für F-Statistiken publiziert.

Jöckel, dessen Dissertation den Forschungsstand über Monte-Carlo Lösungen wiedergibt, warnt ebenfalls vor zu kleinen Umfängen zwischen 500 und 1000 (vgl. schon Marriott, 1979), wie sie häufiger (Edgington & Strain, 1973) in der Literatur verwendet worden sind. Ein Grund liegt darin, daß man auch für 10.000 Zufallsauswahlen schnelle Algorithmen zur Erzeugung von (Pseudo-) Zufallszahlen benötigt. Diese Algorithmen müssen zudem mit hoher Präzision (annähernd) gleichverteilte Zufallszahlen liefern (vgl. die kritischen Anmerkungen von z.B. Afflerbach, 1985; Marsaglia, 1985 und Ripley, 1983). Wenn nämlich z.B. 'zu schnell' Zyklen aufgrund des Bildungsgesetzes der Pseudo-Zufallszahlen auftreten, ist die Voraussetzung der Gleichwahrscheinlichkeit aller möglichen zulässigen Permutationen, in die Zufallsstichprobe aufgenommen zu werden, verletzt. Wegen der Wichtigkeit des Monte-Carlo Verfahrens ist der ganze Abschnitt 8.2 diesem Thema vorbehalten.

(vi) Wenn man nicht an der Bestimmung des exakten p-Wertes selbst oder seiner Schätzung mit dem Monte-Carlo Verfahren interessiert ist, sondern nur eine Testentscheidung treffen will, sind verschiedene Vorgehensweisen prinzipiell möglich. Für eine fest vorgegebene maximale Anzahl auszuführender zufälliger Permutationen und ein vorgegebenes Niveau α wird nach jeder Berechnung des Teststatistikwertes überprüft, ob die Anzahl von 'extremen' (Permutations-) Teststatistikwerten (je nach Fragestellung sind diese größer/gleich oder kleiner/gleich dem ursprünglichen Wert) bereits größer/gleich dem α-ten Teil auszuführender Permutationen ist. Sobald das zum ersten Mal der Fall ist, kann die Nullhypothese nicht mehr abgelehnt werden (vgl. Jöckel, 1982). Überschreitet umgekehrt die Anzahl von 'nicht-extremen'

(Permutations-) Teststatistikwerten (kleiner bzw. größer als der ursprüngliche Wert) zum ersten Mal den (1-α-ten) Teil auszuführender Permutationen, kann die Nullhypothese verworfen werden, da dann der ursprüngliche Teststatistikwert sich nur noch unter dem 'extremen' α-ten Teil aller Teststatistikwerte befinden kann. Besonders wenn keine Abweichungen von der Nullhypothese vorliegen, bedeutet dieses Verfahren eine große Reduktion im Rechenaufwand. Im 'günstigen' Fall kann schon nach wenig mehr als dem α-ten Teil aller geplanten Permutationen das Verfahren beendet werden. Im Fall einer Abweichung von der Nullhypothese ist dagegen mindestens der (1-α)-te Anteil an Permutationen auszuführen. Für beide Testabbruch-Situationen ist auch noch eine Schätzung des p-Wertes möglich (vgl. Jöckel, 1982, S. 71f).

Mit den beiden ersten Momenten der Permutationsverteilung einer Permutationsteststatistik kann man leicht eine u.U. sehr konservative Schätzung des p-Wertes gewinnen (vgl. Oldenbürger, 1981, S. 169). Nach der Ungleichung von Tchebycheff (vgl. Bortz, 1977, S. 56) gilt für jede Zufallsvariable X mit Erwartungswert E(X) und endlicher Varianz Var(X) sowie für jedes positive Vielfache λ der Standardabweichung $(Var(X))^{1/2}$:

(8.1) $P\{|X-E(X)|/(Var(X))^{1/2} \geq \lambda\} \leq 1/\lambda^2$.

Für ein vorgegebenes α-Niveau (z.B. α=.05) muß damit bei zweiseitiger Fragestellung überprüft werden, ob der Wert der standardisierten Teststatistik größer als $(1/\alpha)^{1/2}$ ist. In diesem Fall kann die Nullhypothese verworfen werden. Ansonsten muß der Permutationstest exakt oder nach den zuvor dargestellten verkürzten Verfahren durchgeführt werden. Für den einseitigen Fall ist die Ungleichung von Markoff (vgl.

De Groot, 1975) geeignet. Für jede nichtnegative Zufallsvariable und jedes positive k gilt:

$$P(X \geq k) \leq E(X)/k \quad .$$

Wenn man die Teststatistik so (linear) transformiert, daß der kleinste mögliche Wert gerade nicht negativ wird, kann man bei bekanntem erstem Permutationsmoment der Teststatistik für vorgegebenes α und Realisation s der Teststatistik die Nullhypothese verwerfen, falls

(1) $E(s)/s \leq \alpha$

ist. Ansonsten muß der Permutationstest selbst ausgeführt werden. Für α = .05 z.B. muß allerdings die Realisation der Teststatistik s mindestens 20 mal großer sein als der Erwartungswert. Es ist also nur ein Verwerfen der Nullhypothese bei einer sehr großen Abweichung von ihr zu erwarten.

(vii) Wall & Meystre (1974) haben einen sequentiellen Monte-Carlo Test vorgestellt. Bei einer fest vorgegebenen hinreichend großen Maximalzahl von zufälligen (zulässigen) Permutationen wird nach jeder (oder nach einer festen Anzahl von) Permutation(en) überprüft, ob die Nullhypothese akzeptiert oder verworfen wird, oder ob eine neue (feste Anzahl von) Permutation(en) erzeugt werden muß. Sequentielle Verfahren werden in dieser Arbeit jedoch nicht weiter berücksichtigt.

8.2 Die Monte-Carlo Version des Permutationstests

Monte-Carlo (MC) Methoden bilden eine Klasse von Verfahren, die in vielen Bereichen Anwendung finden (Collier & Larson, 1969). Wenn man ein mathematisches Problem vorliegen hat, das nicht (leicht) durch analytische oder numerische

Methoden lösbar ist, ist es dagegen oft möglich, einen stochastischen (zufälligen) Prozeß zu finden, der Zufallsvariablen mit einer Häufigkeitsverteilung erzeugt, die in 'enger' Beziehung zu der ursprünglichen mathematischen Formel steht (z.B. Hartley, 1977; Rubinstein, 1981). Insbesondere können Monte-Carlo Methoden zur Lösung von statistischen Verteilungsproblemen herangezogen werden. Besonders wichtig ist der Fall, in dem die exakte Verteilung einer Teststatistik nicht bekannt ist bzw. auch mit großen Rechenanlagen nicht mit vertretbarem Aufwand bestimmbar ist. Wie bereits in Abschnitt 8.1 erwähnt, stammt von Dwass (1957) bzw. Barnard (1963) der Vorschlag, MC-Methoden bei der Bestimmung von Permutationsteststatistiken einzusetzen (Pratt & Gibbons, 1981, S. 211ff). Grundlegend ist die Überlegung, daß es mit einem Simulationsexperiment möglich ist, stochastisch unabhängige Realisationen einer Permutationsteststatistik zu gewinnen, die bei Gültigkeit der jeweiligen Nullhypothese alle gleichwahrscheinlich sind.

Mit Hilfe der Darstellung eines Permutationstests als Entscheidungsfunktion wie in Abschnitt 2 (vgl. (2.4) bzw. (2.5)) läßt sich auch die MC-Version angeben. Sei nun $M_o << M$ die feste Anzahl von zufällig ausgewählten Permutationen aus der Menge aller unter der jeweiligen Nullhypothese zulassigen Permutationen der Komponenten des gesamten Beobachtungsvektors.

Seien $x_1,\ldots,x_{M_o}$ die permutierten Beobachtungsvektoren und x die ursprünglichen Beobachtungen selbst. Wie bei Dwass werden die Permutationen mit Zurücklegen gezogen (vgl. Schema 4 in Gabriel und Hall, 1983). Es ist also u.a. möglich, daß unter den $x_1,\ldots x_{M_o}$ der ursprüngliche Beobach-

tungsvektor wiederum auftritt. Je nach Versuchsplan und zu testender Hypothese bezeichnet x bereits die (maximal-) invariant transformierten Beobachtungen.

Dann ist für die Teststatistik S(X) mit Realisation s = S(x) der randomisierte Monte-Carlo (Permutations-) Test zum Umfang M_o und zu erreichbarem Niveau $\alpha = k/(M_o+1)$ definiert durch (vgl. Jöckel, 1982, S. 29):

$$(8.2)\quad \phi(x,x_1,..,x_{M_o}) = \phi(s,s_1,..,s_{M_o}) = \begin{cases} 1 & p^{>}+p^{=} \leq k-1 \\ \gamma(s) = (k-p^{>})/(p^{=}+1) & \text{sonst} \\ 0 & p^{>} \geq k \end{cases}$$

mit

$$p^{>}=\{s_i \mid s_i>s,\ i=1,..,M_o\} \quad \text{und} \quad p^{=}=\{s_i \mid s_i=s,\ i=1,..,M_o\}\ .$$

Bei der praktischen Ausführung des MC-Tests wird i.a. auf die randomisierte Entscheidung mit Randomisierungskonstante $\gamma(s)$ verzichtet und der nichtrandomisierte Monte-Carlo Test

$$(8.3)\quad \tilde{\phi}(x,x_1,..,x_{M_o}) = \tilde{\phi}(s,s_1,..,s_{M_o}) = \begin{cases} 1 & p^{>}+p^{=} \leq k-1 \\ 0 & p^{>}+p^{=} \geq k \end{cases}$$

verwendet. Dieser hat ein Niveau $\leq \alpha$. In den meisten Fällen bedeutet die Wahl von (8.3) keinen großen Unterschied zu (8.2). Wenn man den Teststatistikwert für die ursprünglichen Beobachtungen selbst in $p^{=}$ mitzählt (vgl. die Algorithmen in Abschnitt 8.4), muß k in (8.3) durch k+1 ersetzt werden.

Mann kann nachweisen (vgl. Jöckel, 1982), daß der beobachtete p-Wert des MC-Tests

$$(8.4)\quad p := p(x) = (p^{>}+p^{=})/(M_o+1)$$

ein (stark konsistenter, erwartungstreuer) Schätzwert für den p-Wert des exakten Permutationstests ist. Damit ist

$$(8.5)\quad p \pm u_{1-\alpha/2} \cdot ((p(1-p)/(M_o+1))^{1/2}$$

ein asymptotisches (1-α)-Konfidenzintervall für den exakten p-Wert des exakten Permutationstests. Für $M_o = 9999$, $1-\alpha = 99\%$, und einem MC-p-Wert von p = .05 erstreckt sich das Intervall auf 0.0445 - 0.0557. In Tabelle 8.1 sind exakte bzw. asymptotische 99%-Konfidenzintervalle für verschiedene p-Werte und Simulationsstichprobenumfange angegeben. Eine u.U. recht konservative Abschätzung für alle $p \in (0,1)$ erhalt man, falls man p = .5 betrachtet. Dann ist p(1-p) maximal.

	Simulationsstichprobenumfang M_o+1					
p-Wert	100	500	1000	10000	20000	100000
.001	*	*	+ .0000-.0074	±.0008	±.0006	±.0003
.005	*	+ .0005-.0201	+ .0009-.0139	±.0018	±.0013	±.0006
.01	+ .0001-.0720	+ .0018-.0277	±.0081 .0035-.0211	±.0026	±.0018	±.0008
.025	+ .0002-.1131	±.0180 .0102-.0487	±.0127 .0139-.0405	±.0040	±.0028	±.0013
.05	+ .0109-.1351	±.0251 .0279-.0803	±.0178 .0337-.0703	±.0056 .0445-.0557	±.0040	±.0018
.10	±.0704 .0382-.2020	±.0346 .0681-.1392	±.0245 .0769-.1267	±.0077 .0923-.1078	±.0055	±.0024
.50	±.129 .3983-.6017	±.0577 .4417-.5583	±.0408 .4588-.5412	±.0129 .4871-.5130	±.0091	±.0041

* Niveau ist mit Simulationsstichprobenumfang M_o+1 nicht zu erreichen.

\+ Asymptotisches Intervall nicht sinnvoll, da die untere Grenze negativ würde.

Tab. 8.1.: Asymptotische (1.Zeile) und wenn moglich exakte 99%-Konfidenzintervalle (2.Zeile) für verschiedene p-Werte bei verschiedenen Simulationsstichprobenumfängen der Monte-Carlo Losung fur den Permutationstest. Ab $M_o \geq 16640$ ist das halbe Konfidenzintervall kleiner als .01 selbst für p=0.5. (Exakte Intervallgrenzen aus Geigy-Tabellen, 1969).

Ebenso wichtig wie die Einhaltung des α-Niveaus ist der Nachweis, daß der MC-Test nicht 'wesentlich weniger' Teststärke besitzt als der exakte Permutationstest. Jöckel (1982, Kap. 2.3) konnte nachweisen, daß für jeweils ein festes Niveau α für die Teststärke $\beta_{M_o}(\alpha)$ des MC-Tests für wachsenden Simulationsstichprobenumfang gilt:

$$\beta_{M_o+1}(\alpha) \geq \beta_{M_o}(\alpha) \quad \alpha \in (0,1)$$

(8.6) und

$$\lim_{M_o \to \infty} \beta_{M_o}(\alpha) = \beta(\alpha) \quad \text{(für ein in } \alpha \text{ stetiges } \beta(\alpha)) \; .$$

Die MC-Teststärke nimmt also mit wachsendem Simulationsstichprobenumfang M_o nie ab und konvergiert (gleichmäßig) gegen die Teststärke des exakten Permutationstests. Von Interesse für die praktische Anwendbarkeit des MC-Tests ist nun, wie schnell diese Approximation hinreichend genau ausfällt. Jöckel hat nachgewiesen, daß für die Differenz der Teststärke von exaktem - und MC-Test gilt:

$$(8.7) \qquad \beta(\alpha) - \beta_{M_o}(\alpha) \leq \beta(\alpha) \cdot (1 - e^D_{M_o,\alpha}) \; .$$

Die Größe $e^D_{M_o,\alpha}$ heißt Dwass-Effizienz zum Simulationsstichprobenumfang M_o. Tabelle 8.2 nach Jöckel (1982, S. 52) gibt für verschiedene α-Niveaus an, wie groß M_o mindestens sein muß, damit $e^D_{M_o,\alpha}$ einen vorgegebenen Wert e^D nicht unterschreitet. Die Formel für eine approximative Bestimmung von M_o ist:

$$(8.8) \qquad M_o = (1/2 \cdot \pi)(1-\alpha)/\alpha \cdot (1/(1-e^D)^2) \; .$$

Umgekehrt kann man mit folgender Formel

$$(8.9) \qquad e^D = 1 - ((1-\alpha)/\alpha)^{1/2} \cdot 1/(M_o \cdot 2\pi)^{1/2}$$

approximativ die Dwass-Effizienz für vorgegebenes α und Simulationsstichprobenumfang M_o bestimmen.

e^D	α - Niveau .01	.025	.05	.10
0.7	199	99	39	19
0.8	399	199	79	39
0.9	1599	679	319	149
0.95	6399	2519	1219	579
0.99	157599	62079	30239	14329

Tab. 8.2.: Notwendiger Simulationsstichprobenumfang M_o des Monte-Carlo Tests für eine minimale Dwass-Effizienz e^D bei dem Niveau α (vgl. Jockel, 1982, S. 52)

M_o	α - Niveau .01	.025	.05	.10
99	.601	.750	.825	.880
499	.822	.888	.922	.946
999	.874	.921	.945	.962
9999	.960	.975	.983	.988
19999	.972	.982	.988	.992

Tab. 8.3.: Dwass-Effizienz e^D (approximativ für den Simulationsstichprobenumfang M_o und das Testniveauα .

Wie man aus Tabelle 8.3 entnehmen kann, sind viele der in Abschnitt 6 erwähnten Studien zu Teststärke und Robustheit von Permutationstests mit zu kleinen Simulationsstichprobenumfängen (von M_o = 1000 oder weniger) ausgeführt worden (vgl. Jöckel, 1980).

Aufgrund der Tabellen 8.1 - 8.3 erscheint ein Simulationsstichprobenumfang von mindestens M_o = 9999 erforderlich. Boyett & Shuster (1977) arbeiten mit 17.000 Permutationen, da das halbe asymptotische 99%-Konfidenzintervall für den exakten p-Wert dann stets kleiner als 0.01 ausfällt (Shuster & Boyett, 1979). Exakt ist das ab 16.640 Permutationen der Fall. Edgington (1980) wie auch Recchia & Rocchetti (1982) arbeiten dagegen i.a. mit nur M_o=999. In den eigenen Programmen werden wie bei Matts & McHugh (1983) i.a. 19.999 Permutationen angesetzt; bei Schätzungen der exakten Teststärke dagegen je nach Umfang der Fragestellung 999 oder 9999 (vgl. Abschnitt 8.4).

Im Unterschied zu der Empfehlung von Edgington (1980, S. 43) wird bei den Computer-Ausdrucken der Ergebnisse der Simulationsstichprobenumfang mitangegeben, um dem Benutzer Informationen über die möglichen Abweichungen vom exakten Permutationstest zu geben.

Die Güte dieser Abschätzungen wie auch die Validität des mit dem MC-Test berechneten p-Wertes hängen entscheidend von der Güte des jeweils verwendeten (Pseudo-) Zufallszahlengenerators ab. Dieser muß idealerweise stets gleichverteilte Zufallszahlen generieren, mit deren Hilfe gleichwahrscheinliche zulassige Permutationen der Komponenten des gesamten Beobachtungsvektors gewonnen werden (vgl. Abschnitt 8.3). Nur bei sehr guter Annaherung an die exakte Gleichverteilung ist zu erwarten, daß beim MC-Test nicht identische Permuta-

tionen wiederholt auftreten bzw. Vertauschungen zwischen bestimmten Faktorstufen mehr oder weniger wahrscheinlich sind. Marsaglia (1985) gibt einen aktuellen Überblick über Stärken und Schwächen einer Vielzahl publizierter Pseudo-Zufallszahlengeneratoren (vgl. auch Afflerbach (1985) für Mikro-Computer). Er schlägt auch acht verschiedene 'stringente' Tests auf Gleichverteilung vor, die besonders für Generatoren, die mehrere einfachere nicht perfekte Generatoren kombinieren, stets befriedigende Ergebnisse liefern und eine sehr große Periode aufweisen, d.h. erst nach sehr vielen Permutationen in denselben Zyklus von Permutationen geraten. In den eigenen Programmen wird der maschineninterne (lineare Kongruenz-) Generator (RAN) des verwendeten Rechners VAX 11/780 benutzt. Dessen Eigenschaften in den acht Tests von Marsaglia sind nicht bekannt. Ein weiteres Kriterium für die Güte des Generators ist die für die Erzeugung einer Zufallszahl benötigte Zeit. In einem kleinen Vergleich mit dem als sehr exakt beschriebenen Generator von Wichmann & Hill (1982) benötigt der VAX-eigene Generator nur ca. 1/6 der Zeit (3,08 sec gegenüber 0,55 sec für 20000 Zufallszahlen). Einen sehr groben Test der Güte eines Generators kann man durchführen (vgl. auch Edgington, 1980, S. 42ff), indem man für ein Testproblem das eine exakte p-Wert Bestimmung mit exakt M_o Permutationen gerade noch zulaßt, wiederholt MC-Tests mit derselben Anzahl M_o von zufälligen Permutationen ausführt. Anschließend überprüft man, ob sich nur ungefähr der erlaubte Anteil von 1% aller berechneten MC-p-Werte außerhalb des 99%-Konfidenzintervalls um den exaken p-Wert herum ergibt. Für eine Reanalyse der artifiziellen Daten von Wilkinson (1975) in einem 3-dimensionalen unver-

bundenen 3-Stichprobenproblem mit je 5 Beobachtungen je Faktorstufe mit exakt 126.126 zulässigen Permutationen und einem exakten p-Wert von p=.01286 für das Pilai'sche Spurkriterium (vgl. (4.18) und (4.29)), sind bei 100 MC-Tests mit je M_o = 19999 Permutationen die extremen MP-p-Werte p_{min} = .0110 und p_{max} = .01465 alle innerhalb des 99%-Konfidenzintervalles von .0108 - .0149 um den exakten p-Wert.
Abschließend sei noch darauf hingewiesen, daß das Ziehen mit Zurücklegen in dem obigen Beispiel wesentlich weniger rechenzeitaufwendig ist als Ziehen ohne Zurücklegen. Im letzteren Fall müßte man über alle bisherigen zufälligen Permutationen der Beobachtungen Buch führen. Das würde neben viel Speicherplatz auch viel Rechenzeit für den Vergleich jeder neuen Permutation mit allen bisher ausgeführten Permutationen erfordern. Falls eine neu erstellte Permutation schon in der Liste bisheriger Permutationen enthalten wäre, müßte sie vernachlässigt werden. Bei guten Pseudo-Zufallszahlengeneratoren und M_o viel kleiner als die exakte Anzahl zulässiger Permutationen ist der mögliche Unterschied im p-Wert sicher praktisch unerheblich.

8.3 Allgemeiner Aufbau von (Invarianz)-Permutationstestprogrammen

Alle Programme für Randomisierungs- oder Permutationstests, die explizit Permutationen der Komponenten des gesamten Beobachtungsvektors beinhalten, sind nach demselben generellen Schema angelegt. Dieses Schema ist wohl auch wegen seiner allgemeinen Bedeutung auf der Umschlagseite des Buches von Edgington (1980) wiedergegeben. Unter der nicht

einschränkenden Bedingung, daß 'große' Werte der Teststatistik gegen die jeweilige Nullhypothese sprechen sollen, hat dieses Flußdiagramm folgendes Aussehen.

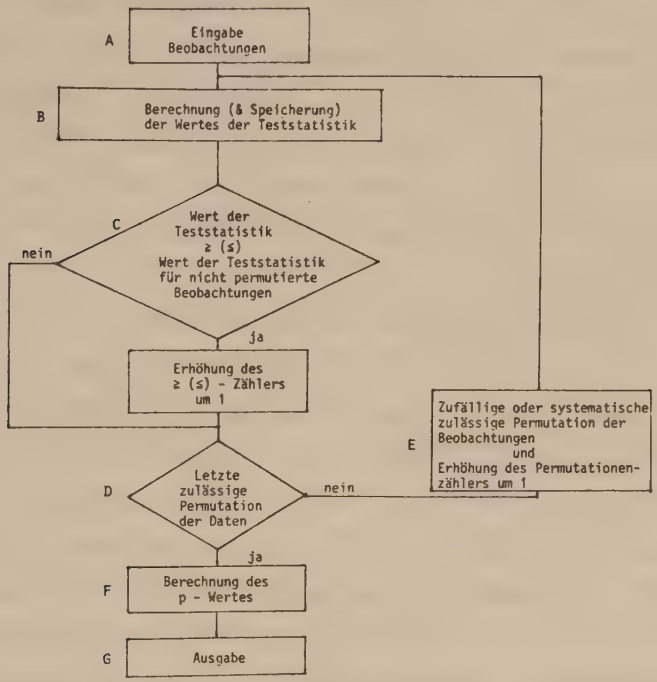


Abb.8.1: Allgemeines Flußdiagramm für einen Permutationstest

Je nach Testproblem sind an den Stellen A - G des allgemeinen Flußdiagramms Spezifizierungen vorzunehmen.

A: Vor und zusätzlich zu der Eingabe der Beobachtungen selbst sind Parameter des verwendeten Versuchsplans wie Anzahl der Faktoren, Faktoren mit/ohne Meßwiederholungen, uni/multivariate Beobachtungen, Anzahl Faktorstufen, Anzahl Replikationen je Faktorstufe(nkombination) zu spezifizieren. Oft ist es auch sinnvoll, vor Bearbeitung von B vorab festzulegen, ob ein exakter Permutationstest oder eine Monte-Carlo Lösung mit einer vorzugebenden Anzahl von zufällig

ausgewählten Permutationen auszuführen ist. Um dem Benutzer nicht die Bestimmung der Anzahl zulässiger Permutationen $|\Omega_n|$ zuzumuten, kann - je nach zu testender Hypothese u.U. verschieden - eine Maximalzahl auszuführender Permutationen - etwa (NPMAX=) 20.000 - vorgegeben werden. Eine Programmroutine kann dann anhand der eingegebenen Versuchsplancharakteristika die exakte Permutationsanzahl bestimmen. Sobald diese die Grenze NPMAX übersteigt, wird stets eine Monte Carlo Lösung mit eben diesem Simulationsstichprobenumfang gewählt.

B: Vor Berechnung des Wertes der Teststatistik ist für viele Testprobleme bei Ausführung eines Invarianz-Permutationstests zuerst eine (maximal-)invariante Transformation der Beobachtungswerte auszuführen (vgl. Abschnitt 3.1 und 4.3). Es kann auch sinnvoll sein, solche Teile der Teststatistik, die durch zulässige Permutationen nicht verändert werden, getrennt von permutationsabhängigen Teilen zu berechnen. Dadurch kann u.U. erheblich an Rechenzeit gespart und an Rechengenauigkeit gewonnen werden. U.U. ist es auch notwendig, aus einer Menge alternativer Teststatistiken für dasselbe Testproblem auszuwählen. Z.B. für multivariate Permutationstests stehen ja verschiedene Teststatistiken mit unterschiedlichen Eigenschaften zur Verfügung. Immer ist bei der Programmierung der Berechnung des Teststatistikwertes darauf zu achten, daß wegen der Vielzahl gleichartiger Berechnungen möglichst effizient und auch numerisch genau zu rechnen ist. So ist z.B. offensichtlich, daß das Pillai'sche Spurkriterium wesentlich weniger Zeit zur Berechnung erfordern muß als alle Kriterien mit expliziter Eigenwertberechnung, da die Bildung der Spur einer Matrix (Summe der

Diagonalelemente) weniger aufwendig ist als die Eigenwertberechnung (vgl. Abschnitt 8.4). Für bestimmte Trendtests aus Abschnitt 7.4 ist in B auch die Bildung einer isotonen Regression z.B. mit dem Algorithmus von Cran (1980) erforderlich.

Für bestimmte multiple Testprozeduren (vgl. Abschnitt 5.3) sind die Teststatistikwerte unter allen ausgeführten Permutationen auch zu speichern.

C: Wenn als Beobachtungen sogenannte Gleitkommazahlen und Divisionen in den Formeln zur Berechnung der Teststatistik auftreten, ist mit Rundungsfehlern in nicht unbeträchtlichem Umfang zu rechnen. Ähnlich verhält es sich, wenn zur Berechnung bestimmter Teststatistiken im multivariaten Fall explizit Eigenwerte bestimmt werden müssen bzw. eine verallgemeinerte Inverse zu bilden ist. In diesen Fällen kann die größer/gleich-Abfrage fehlerhafte Ergebnisse liefern. Für univariate Versuchspläne hatte Pyhel (ohne Jahr) deshalb durchweg mit divisionsfreien Ganzzahl-Algorithmen gearbeitet (vgl. Abschnitt 8.4.1). Dieses Vorgehen findet man auch bei Pagano & Tritchler (1983), Streitberg & Röhmel (1983, 1986) und Zimmermann (1985a, b). Die Ganzzahl-Algorithmen stoßen wegen der veränderten Teststatistiken oft an die obere Grenze der größten im Computer darstellbaren ganzen Zahl (Integer-Zahl). Einerseits müssen Rohdaten mit k Dezimalstellen durch Multiplikation mit 10^k ganzzahlig gemacht und Divisionen durch Stichprobenumfänge oder Faktorstufenanzahlen durch entsprechende Multiplikationen ersetzt werden. Beispielsweise erhält man im Split-Plot-Plan die zu (3.66) äquivalente Teststatistik:

$$(8.10)\quad S^{*}(Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} \left(\prod_{l \neq i}^{r} n_l \right) \cdot n_{\cdot}^{2} c^{2} T_{i.j}^{2}$$

mit

$$n_{\cdot} cT_{ih(i)j} = n_{\cdot} cY_{ih(i)j} - n_{\cdot} Y_{ih(i).} - cY_{..j} + Y_{...} \quad .$$

Aus Formel (8.10) ist sofort ersichtlich, daß selbst für relativ kleine Stichprobenumfänge und Faktorstufenanzahlen es zum sog. 'Integer-overflow' im Rechner kommen kann. Deshalb bleibt besonders für multivariate Testprobleme nur folgender Ausweg:

(i) Alle vorkommenden Formeln sind so zu programmieren, daß möglichst geringe Rundungsfehler bei Verwendung von Gleitkommazahlen entstehen (z.B. Divisionen nur außerhalb von DO-Schleifen etc.).

(ii) Mit Hilfe der sogenannten relativen Maschinengenauigkeit (vgl. z.B. Pagano & Tritchler, 1983) muß möglichst gut für Beobachtungen mit einer spezifizierten Anzahl von Dezimalstellen abgeschätzt werden (vgl. Abschnitt 8.4.), wie weit Teststatistikwerte nach Datenpermutationen verschieden ausfallen können, obwohl ohne Rundungsfehler identische Werte resultieren müßten. Dann sind alle Teststatistikwerte, die innerhalb von $\pm\varepsilon$ (z.B. $\varepsilon=10^{-4}$) um den ursprünglichen Teststatistikwert herum liegen, unter $p^{=}$ (vgl. (3.18)) zu berücksichtigen.

D. E: Da im allgemeinen eine große Zahl von Permutationen der (transformierten) Beobachtungen ausgeführt werden muß, sind schnelle Permutationsalgorithmen von großer Bedeutung. Diese müssen je nach Versuchsplan mit unabhängigen/abhängigen Beobachtungen verschieden beschaffen sein (vgl. z.B. Nijenhuis & Wilf, 1975; Edgington, 1980; Pyhel, ohne Jahr). Einen Überblick und einen schnellen Algorithmus für unverbundene Probleme gibt Berry (1982). Pyhel benutzt nur Algorithmen, in denen Permutationen so weit wie möglich

lediglich durch wenig aufwendige Inversionen (d.h. Vertauschungen von je zwei Elementen) auseinander hervorgehen (vgl. Abschnitt 8.4.1.). Je nach Algorithmus wird entweder $|Q_n|$-mal bzw. NPMAX-mal die Schleife B-C-D-E-B nach einmaligem Anruf in einem Unterprogramm durchlaufen (Berry, 1982) oder - solange die aktuelle Anzahl ausgeführter Permutationen nicht größer als die Gesamtzahl an Permutationen ist - $|Q_n|$ - bzw. NPMAX-mal ein Unterprogramm zur Generierung einer neuen Permutation aufgerufen (z.B. Edgington, 1980; Pyhel, ohne Jahr). Für die Abfrage in D muß natürlich ein Permutationenzähler bei jedem Durchgang durch E um Eins erhöht werden.

Fur die Monte-Carlo Version sei nochmals die Wichtigkeit eines Pseudo-Zufallszahlen Generators erwähnt, der mit hoher Genauigkeit und langer Periode im Intervall (0,1) gleichverteilte Zufallszahlen erzeugt. Durch folgende kurze Abfolge von FORTRAN-Befehlen (vgl. Algorithmus RANPER, Pyhel, ohne Jahr) erhält man daraus bei jedem Aufruf dieser Sequenz durch N Inversionen eine (neue) Permutation der Zahlen 1,..,N im Feld Z(1),..,Z(N):

```
      C ***  Initialisierung
             DO I=1,N
             Z(I)=I
             END DO
      C
             DO I=1,N
      C
(8.11) C ***  RAN( ) liefert auf (0,1)
       C ***  gleichverteilte Zufallszahlen
       C
             L=I+RAN(K1)*(N+1-I)
       C ***  Inversion
             K=Z(L)
             Z(L)=Z(I)
             Z(I)=K
             END DO
```

Die Anzahl an Permutationen im Monte-Carlo Fall muß sich insbesondere nach der angestrebten Teststärke des MC-Tests richten. Weiterhin ist die zu erwartende Größe des p-Wertes von Bedeutung (vgl. Tabelle 8.2 und 8.3). Aus der Struktur von (8.11) ist ersichtlich, daß es sich um 'Ziehen mit Zurücklegen' handelt.

F: Je nach Art der zu testenden Hypothese und nach Art der gewünschten Testentscheidung sind folgende Bestimmungen der (deskriptiven) Statistik p-Wert zu unterscheiden. Falls eine randomisierte Testentscheidung zu fällen ist (vgl. (2.5) und (8.2)), muß nicht nur der p-Wert aus $p^{>}$ und $p^{=}$ (vgl. (3.18)) bestimmt werden. Vielmehr muß u.U. $\gamma=(\alpha-p^{>})/p^{=}$ berechnet und nach Ziehen einer auf dem Intervall (0,1) gleichverteilten Zufallszahl u folgendermaßen entschieden werden:

Die Nullhypothese wird verworfen, falls $u \geq \gamma$ ausfällt.

Es ist auch möglich, wie z.B. bei König, Giani & Haux (1986) einen sog. randomisierten p-Wert zu bestimmen über:

(8.12) $p=p^{>}+up^{=}$.

Ein weiteres Problem ist die Bestimmung des p-Wertes bei einem zweiseitigen Test (vgl. z.B. die Diskussion in Gibbons & Pratt, 1975; Pratt & Gibbons, 1981, S. 28ff). Die gängigste Lösung ist, neben der Angabe der beiden einseitigen p-Werte $p^{\geq}$ und $p^{\leq}$ selbst bei zweiseitigen Problemen folgende Definition eines zweiseitigen p-Wertes:

(8.13) $p=2\min\{p^{\geq},p^{\leq},0.5\}$.

Diese Definition wird auch im (un)verbundenen Zweistichprobenproblem in den eigenen Programmen verwendet.

Falls es sich um einen MC-Test handelt, kann man nach der Entscheidung, ob H_0 zu verwerfen oder beizubehalten ist, noch über (8.5) ein (asymptotisches) Konfidenzintervall für

das exakte Niveau des exakten Permutationstests berechnen bzw. aus Tabelle 8.1 entnehmen.

Auch die Inversion von Permutationstests, wie in Abschnitt 6.2.4 dargestellt, läßt sich am Flußdiagramm aus Abb. 8.1 veranschaulichen. Die Berechnung des Teststatistikwertes in B erfolgt nach Formel (6.10). In C muß die Abfrage durch 'kleiner/gleich Δ ' wie in (6.11) ersetzt und eventuell der 'kleiner/gleich Zähler' erhöht werden. Falls die gesamte p-Wert Kurve in Abhängigkeit vom Lageparameter Δ zu ermitteln ist, fallen die Abfrage C sowie die Zählererhöhung weg und alle Werte der Anker-Statistik $T_\pi^+(y)$ müssen gespeichert werden. Anstelle von F ist nun nach Ordnen aller T_π^+ -Werte nach aufsteigender Größe die $p^+(\Delta)$-Treppenfunktion tabellarisch darzustellen bzw. mit Hilfe eines Plotters zu zeichnen.

Wenn man nicht an der Bestimmung des p-Wertes, sondern im wesentlichen nur an einer Testentscheidung zum Niveau α interessiert ist, kann ein abkürzendes Verfahren verwendet werden (Jöckel, 1982, vgl. Abschnitt 8.1 (vi)). Anstelle des Flußdiagramms in Abb. 8.1 hat man dann folgenden Ablauf wie in Abb. 8.2, der sowohl für die MC-Version wie den exakten Test gültig ist:

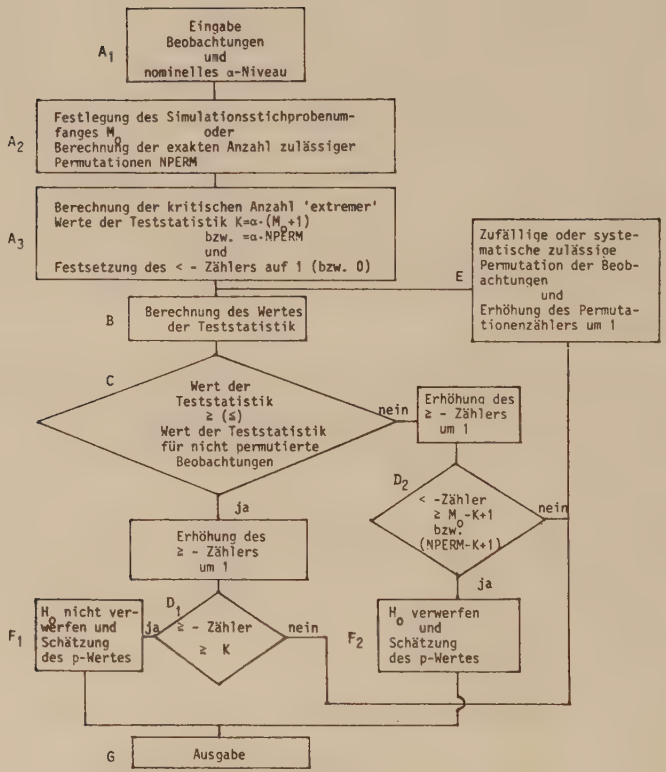


Abb. 8.2: Allgemeines Flußdiagramm für einen Permutationstest ohne Bestimmung des p-Wertes sondern nur für die Entscheidung 'H_0 verwerfen' oder 'H_0 nicht verwerfen'. In Klammern sind die Optionen für die Monte-Carlo Version angegeben.

Die Bezeichnungen zu aus Abb. 8.1 strukturell identischen bzw. ähnlichen Komponenten wurden in Abb. 8.2 mit denselben Buchstaben vorgenommen. In A_3 heißt es M_o+1 bei der Berechnung von K, aber NPERM für den exakten Test, da in der MC-Version der ursprüngliche Teststatistikwert als 'erste zufällige' Auswahl mitzuzählen ist. Entsprechend ist der Zähler für die kleiner als der ursprüngliche Teststatistikwert ausfallenden 'permutierten Teststatistikwerte in der MC-Version auch bereits mit 1 (nicht mehr die Zahl 0) vorzubesetzen. Zur Entscheidung für eine der beiden Alternativen 'Nullhypothese nicht verwerfen' bzw. 'Nullhypothese verwerfen' werden zwei Zähler ('größer/gleich' bzw. ' kleiner') benötigt. Es werden solange in B Teststatistikwerte mit in E zufällig (bzw. systematisch) erzeugten Permutationen der Beobachtungen berechnet, bis einer der beiden Zähler in D_1 bzw. D_2 die kritische Anzahl K bzw. M_o - K+1 für ein Beibehalten oder Verwerfen der Nullhypothese erreicht. Falls der Größer/Gleich-Zähler K erreicht ist, ist der ursprüngliche Wert der Teststatistik nicht unter den 'extremen' Realisationen. Falls der Kleiner-Zähler M_o - K+1 erreicht, muß der ursprüngliche Wert in der kritischen Region liegen. Als Schätzung des exakten p-Wertes (vgl. Jöckel, 1982, S. 72) in F_1 bzw. F_2 erhält man, wenn l der aktuelle Wert des Permutationenzählers in E ist und der Algorithmus über F_1 bzw. F_2 beendet wird:

(8.14) $\hat{p}=K/l$ bzw. $\hat{p}=(l-(M_o-K+1))/l$;

Im exakten Fall ist bei F_2:

p=(l-(NPERM-K+1))/l .

Zur exakten Bestimmung sowie zur Schätzung der exakten Teststärke (vgl. Abschnitt 6.3) des Permutationstests gegen eine feste (Lage-)Alternative für ein vorgegebenes α-Niveau ist

wiederum das Flußdiagramm in Abb. 8.1 grundlegend. Abb. 8.3 zeigt, daß zur exakten Bestimmung der exakten konditionalen Teststärke das Flußdiagramm aus Abb. 8.1 als Element in ein zweites strukturgleiches Schema eingebettet worden ist. Je nach Testproblem mussen die Beobachtungen in A u.U. zuerst (maximal-) invariant transformiert werden.

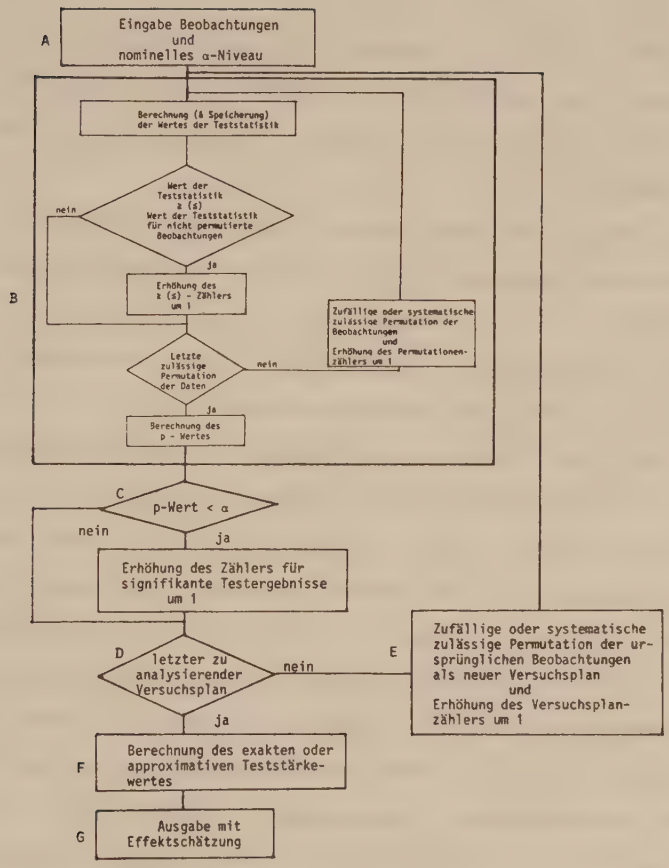


Abb. 8.3: Allgemeines Flußdiagramm für die Bestimmung der exakten oder approximativen Teststarke des Permutationstests zum nominellen Niveau α.

Wegen des zu großen Rechenaufwandes für eine exakte Lösung ist nur eine zweifache MC-Lösung realistisch ausführbar. Die Genauigkeit der einzelnen Berechnungen des p-Wertes innerhalb von B sollte größer sein als die anschließende Bestimmung der Teststärke $\hat{q}^{>}$ in F, wenn man der Argumentation in Abschnitt 6.2.3 bzw. am Ende von Abschnitt 6.3.1 folgt. 1000 'äußere' zufällige Versuchspläne mit $M_o+1 = 10.000$ bis 20.000 'inneren' zulässigen Datenpermutationen sind realistisch. Eine konservative Angabe für ein 99%-Konfidenzintervall für $q^{>}$ kann man wieder Tabelle 8.1 für $q^{>}(\hat{=}p) = 0.5$ entnehmen.

Die Bestimmung einer ganzen Teststärkekurve für das (un-)verbundene Zweistichprobenproblem in Abhängigkeit vom Lageparameter, wie in Abschnitt 6.2.4 beschrieben, kann anhand des Flußdiagramms aus Abbildung 8.3 veranschaulicht werden. Die Berechnung der Teststatistikwerte erfolgt nach Formel (6.10). Bei Bestimmung der gesamten Teststatistikkurve für ein eingegebenes Niveau α entfällt in dem B-Komplex der Vergleich des Teststatistikwertes mit Δ und die eventuelle Erhöhung des 'kleiner/gleich'-Zählers. Vielmehr müssen wieder alle $|\Omega_n|^{-1}$ Anker-Statistikwerte (d.h. eine Spalte in Tabelle 6.9) gespeichert und statt der p-Wert Berechnung das erreichbare untere $|\Omega_n|_\alpha$ -Quantil dieser Anker-Statistikwerte bestimmt werden. Entsprechend entfällt die Abfrage in C mit der anschließenden Erhöhung des 'signifikante Testergebnisse'-Zählers. Statt F muß nach Betrachtung aller zu analysierenden Versuchspläne die empirische Verteilungsfunktion der negativen Werte der obigen Quantile tabellarisch oder mit einem Plotter graphisch dargestellt werden. Erneut ist i.a. eine doppelte MC-Lösung notwendig. Wenn man nur an der Teststärke für ein Niveau α interessiert ist, kann die

Speicherung der ganzen Werte-Matrix wie in Tabelle 6.2 entfallen. Dann genügen eine Spalte und ein Feld, das die einzelnen Quantilwerte enthält.

Zur naiven Approximation der finiten konditionalen Teststärke müssen, wie in Abbildung 8.4 dargestellt, zwei strukturgleiche Flußdiagramme aus Abbildung 8.1 hintereinander angeordnet werden.

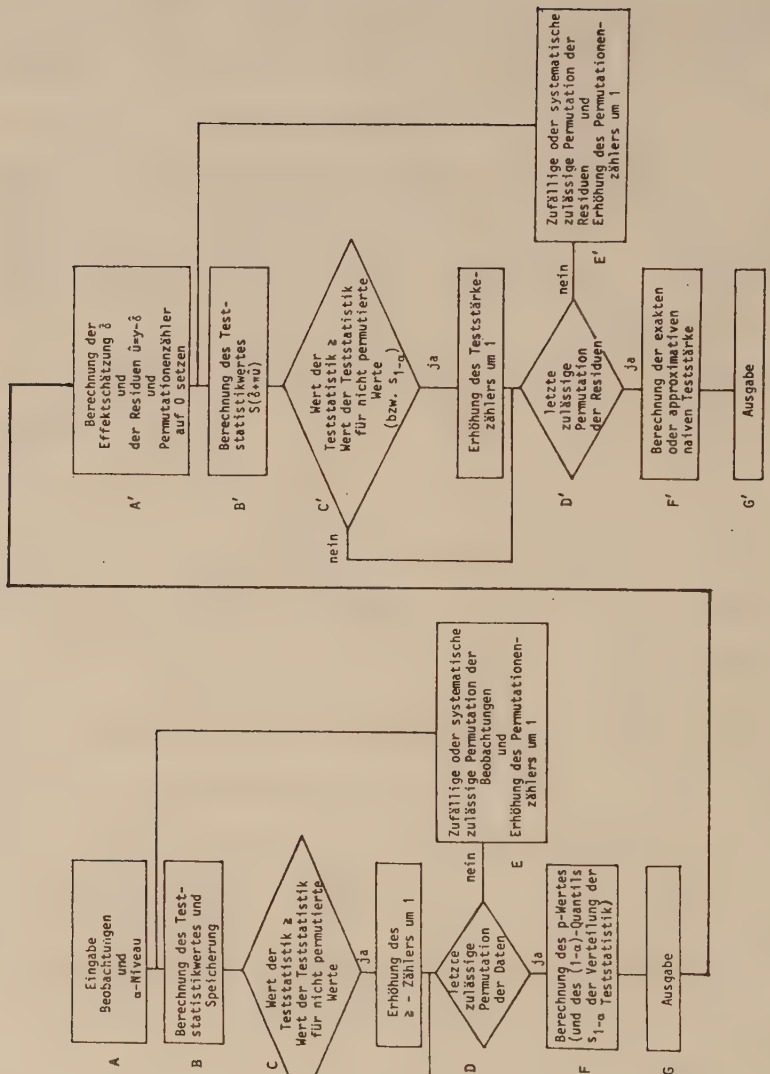


Abb. 8.4: Allgemeines Flußdiagramm für die 'naive' Bestimmung der Teststärke eines Permutationstests.

Je nachdem, ob man die naive (approximative) Teststärke zum beobachteten Testniveau ausgedrückt durch den p-Wert oder zum erreichbaren Niveau α bestimmen will, muß in F lediglich der p-Wert berechnet werden. Ansonsten sind in B alle Teststatistikwerte zu speichern, und es ist zusätzlich in F das $(1-\alpha)$-Quantil dieser Teststatistikwerte zu bestimmen. Falls aufgrund von Bindungen in den Teststatistikwerten ein gewünschtes Niveau α nicht erreichbar ist, wird das nächste größere erreichbare Niveau gewählt.
Anstelle des beobachteten Effektes $\hat{\delta}$ kann auch die Teststärke gegen einen hypothetisch angenommenen Effekt δ_0 ermittelt werden. Dann ist dieser in A' einzugeben und in B' der Wert $S(\delta_0+\pi(\underline{y}-\delta_0))$ zu berechnen.

Eine konzeptuell andere Herangehensweise an das Problem, explizit die Permutationsverteilung einer Permutationsteststatistik zu bestimmen, stellen sogenannte Verschiebe-Algorithmen (Shift Algorithmen, Neumann, 1985; Streitberg & Röhmel, 1986; Oldenbürger, 1981; Zimmermann, 1985a, b) dar. Sie sind allerdings nur dann verwendbar, wenn sich die Permutationsteststatistik eventuell nach geeigneten Umformungen als Linearkombination von unabhängigen (i.a. identisch verteilten) Zufallsvariablen ausdrücken läßt. Ohne Einschränkung der Allgemeinheit wird stets angenommen, daß die Beobachtungen ganzzahlig und eventuell zusatzlich nichtnegativ sind. Das ist stets durch geeignete Transformationen der Daten, die alle das Testproblem unverändert lassen, erreichbar.
Die grundlegende Idee (Streitberg & Röhmel, 1986; Zimmermann, 1985a, b) läßt sich am leichtesten am verbundenen Zweistichprobenproblem (vgl. Abschnitt 3.1.5 (ii)) ver-

anschaulichen. Es sei vorausgesetzt, daß die Beobachtungen ganzzahlig sind.

Die Teststatistik für den einseitigen Test (3.39) ist äquivalent zu einer Teststatistik, in der pro Beobachtungseinheit (Block) i das Minimum $C_i = \min(Y_{1i}, Y_{2i})$ subtrahiert worden ist. Sei $Y^*_{ji} = Y_{ji} - C_i$, $j=1,2$; $i=1,..,n$. Damit erhält man pro Block transformierte Beobachtungen, die sich um $A_i = |Y_{1i} - Y_{2i}|$ unterscheiden und deren kleinere stets Null ist.

Unter der Nullhypothese der Permutationsinvarianz der Verteilung von y^* unter den 2^n Intra-Blockpermutationen ist dann die Verteilung von

(8.15) $$S(Y^*) = \sum_{i=1}^{n} Y^*_{1i}$$

identisch mit der bedingten Verteilung von

(8.16) $$T(Y^*) = \sum_{i=1}^{n} a_i \cdot K_i \quad , \; a_i = |y_{1i} - y_{2i}| \; , \; i=1,..,n,$$

wobei die K_i n unabhängig und identisch Bernoulliverteilte Zufallsvariablen mit $P(K_i=1)=P(K_i=0)=0.5$ sind.

$T(Y^*)$ hat gegenüber $S(Y)$ den Vorteil, daß man dessen Verteilung besonders leicht ohne explizite Permutation der Beobachtungen je Block ermitteln kann. Wegen der Unabhängigkeit der K_i läßt sich die diskrete Wahrscheinlichkeitsverteilung von $T(Y^*)$ unter (n-1)-facher Anwendung des sog. Faltungssatzes für zwei unabhängige diskrete Zufallsvariablen (Meschkowski, 1968, S. 140ff) sukzessiv aufbauen (Oldenbürger, 1981). Wegen der besonderen Form der Y^*_{ji} - für jedes i ist einer der beiden Werte identisch Null - ist diese Vorgehensweise äquivalent mit der n-fachen Ausführung der folgenden Rekursionsformel für die Häufigkeit $h_n(t), t=0,..,\sum_{i=1}^{n} a_i$, mit der nach n Schritten die möglichen

Werte $t=0,\ldots,\sum_{i=1}^{n} a_i$ der Teststatistik $T(Y^*)$ bei transformierten Beobachtungen y^* auftreten. Es gilt:

(8.17) $h_k(t)=h_{k-1}(t)+h_{k-1}(t-a_k)$, $k=1,2,\ldots,n$.

mit der Rekursionsverankerung $h_o(t)=\begin{cases}1 & t=0\\ 0 & t\neq 0\end{cases}$

und

$h_k(t)=2^k P(T(Y^*)=t)$.

Dabei ist $1/2^k$ die Wahrscheinlichkeit für jede der zulässigen Permutationen innerhalb der k Beobachtungseinheiten.
Die Wirkungsweise der Verschiebe-Operationen um a_k in (8.17) verdeutlicht man sich leicht an einem Beispiel.
Subtrahiert man für die 5. abhängige Variable in Tabelle 5.1 das jeweilige Minimum von Vor- und Nachtestwert, erhält man die Ergebnisse aus Tabelle 8.4:

Pat.	Original Daten		Minimum	Transform. Daten		Wert
i	vor	nach		vor	nach	a_i
1	8	5	5	3	0	3
2	6	7	6	0	1	1
3	13	12	12	1	0	1
4	14	19	14	0	5	5
n=5	13	14	13	0	1	1
Summe	54	57	50	4	7	11

Tab. 8.4: Beispiel: Aufbereitung der Rohdaten für den Verschiebealgorithmus im verbundenen Zweistichprobenproblem; Originaldaten sind die Werte für die 5. abhängige Variable aus Tabelle 5.1.

Außerdem sind die absoluten Differenzen a_i aus Vor- und Nachtestwert in der Tabelle angegeben. Minimaler und maximaler Wert in der Permutationsverteilung der Teststatistik (8.15) sind damit 0 und $\sum_{i=1}^{n} a_i=11$. Die Ausführung der fünf Verschiebeoperationen zeigt Tabelle 8.5:

$h(\sum_{i=1}^{k} y^*_{1i} = t)$		0	1	2	3	4	5	6	7	8	9	10	11	
k=0:		1												
k=1:		[0	0	0]	1									(+)
	(1)	1	0	0	1									
k=2:		[0]	1	0	0	1								(+)
	(2)	1	1	0	1	1								
k=3:		[0]	1	1	0	1	1							(+)
	(3)	1	2	1	1	2	1							
k=4:		[0	0	0	0	0]	1	2	1	1	2	1		(+)
	(4)	1	2	1	1	2	2	2	1	1	2	1		
k=5:		[0]	1	2	1	1	2	2	2	1	1	2	1	(+)
	(5)	1	3	3	2	3	4	4	3	2	3	3	1	

Tab. 8.5: Beispiel: Verschiebealgorithmus für das verbundene Zweistichprobenproblem mit den Beispieldaten aus Tabelle 8.4.

Man erkennt leicht, daß in den jeweiligen Summenzeilen (1),...,(5) die absoluten Häufigkeiten, mit denen die verschiedenen Realisationen t der 'partiellen' Teststatistik $S(Y^*_{(k)}) = \sum_{i=1}^{(k)} Y^*_{1i}$ auftreten, wiedergegeben sind. Addiert man zu den verschiedenen Realisationen von (8.16) jeweils 50, die Summe der Blockminima aus Tabelle 8.4, so erhält man die geordneten Statistikwerte aus Tabelle 5.3. Man muß jeden Wert lediglich so oft anführen, wie in der letzten Zeile in Tabelle 8.5 angegeben ist.

Der p-Wert für die tatsächlich beobachtete Summe der transformierten Nachtestwerte von 7 (bzw. 57 in den ursprünglichen Werten) ist

$$p = 1/32 \sum_{t=7}^{11} h_5(t) = 1/32(3+2+3+3+1) = 12/32$$

wie in Tabelle 5.3.

In Abhängigkeit von den absoluten Differenzen a_i und deren Anzahl n kann das Feld $h(0),..,h(\sum_{i=1}^{n} a_i)$ u.U. sehr groß ausfallen, sodaß es zu Speicherplatzproblemen kommen kann. Diese bilden die einzige Beschrankung fur den Algorithmus. Identische Werte je Block - d.h. $a_i=0$- stellen kein Problem dar. Sie führen zu einer Verschiebung um 0 Stellen nach rechts.

Man erkennt weiterhin, daß die Verteilung der Häufigkeiten symmetrisch ist, sodaß explizit nur die Hälfte des Feldes in der letzten Zeile bei gerader Anzahl $1+\sum_{i=1}^{n} a_i$ bzw. $1+[(1+\sum_{i=1}^{n} a_i)/2]$ bei ungerader Anzahl berechnet werden muß (vgl. Zimmermann, 1985a).

Im unverbundenen Zweistichprobenproblem mit der Teststatistik (3.38) muß man einen komplexeren 'zweidimensionalen' Verschiebe-Algorithmus (Streitberg & Röhmel, 1986) anwenden. Wiederum seien ganzzahlige und nicht-negative Werte vorausgesetzt. Letzteres erreicht man stets durch Subtraktion des Minimums $c=\min\{Y_1,..,Y_{n_1}, Y_{n_1+1},..,Y_{n_1+n_2}\}$ von allen Beobachtungen $Y^*_i=Y_i-c$, $i=1,..,n_1+n_2 = n$. Unter der Nullhypothese der Permutationsinvarianz der Verteilung von $Y^* = (Y_1,..,Y_n)'$ unter allen n! Vertauschungen der Komponenten ist die Permutationsverteilung der Summe der ersten n_1 Komponenten von Y^*:

(8.18) $$S(Y^*) = \sum_{i=1}^{n_1} Y^*_i$$

identisch mit der bedingten Verteilung der Summe (über alle $n_1 + n_2 = n$):

(8.19) $$T(y^*) = \sum_{i=1}^{n} y^*_i \cdot K_i$$

mit n identisch Bernoulli-verteilten Zufallsvariablen K_i mit $P(K_i = 1) = P(K_i = 0) = 1/2$.

Entscheidend ist, daß im Gegensatz zu (8.16) noch folgende Bedingung an die K_i geknüpft ist:

(8.20) $$n_1 = \sum_{i=1}^{n} K_i \quad .$$

Da der obige Verschiebealgorithmus jedoch nur für unabhängige und identisch verteilte Zufallsvariablen gilt, muß man zusätzlich folgendes überlegen:

Auf die gemeinsame Verteilung von

(8.21) und
$$T(y^*) = \sum_{i=1}^{n} y_i^* \cdot K_i$$
$$N_1 = \sum_{i=1}^{n} 1 \cdot K_i$$

läßt sich dagegen eine zweidimensionale Version des Verschiebealgorithmus aus (8.17) anwenden:

(8.22) Sei $A=((a_{1i}))_{\substack{1=1;..;k\\ i=1;..;n}}$ eine Matrix von Gewichten $=(a_1,..,a_n)$ mit Spaltenvektoren $a_i=(a_{1i},..,a_{ki})'$,

$K=(K_1,..,K_i,..,K_n)'$ ein Vektor von unabhängig identisch verteilten Bernoulli-Zufallsvariablen mit

$P(K_i=1)=P(K_i=0)=1/2,\ i=1,..,n$

und $T=(T_1,..,T_k)'$ ein Vektor von Teststatistiken

mit
$$T = A\cdot K = (\sum_{i=1}^{n} a_{1i}K_1, \ldots, \sum_{i=1}^{n} a_{1i}K_1, .., \sum_{i=1}^{n} a_{ki}K_k)' \quad .$$

Wegen der Unabhängigkeit der K_i läßt sich die k-dimensionale Verteilung der $T_1,..T_k$ wiederum nach dem Faltungssatz sukzessive aufbauen. Die Häufigkeit, mit der die verschiedenen Werte von T auftreten, läßt sich durch n-fache Ausführung der folgenden Rekursionsformel bestimmen, wobei h jetzt eine vektorwertige Funktion ist:

$$h_\nu(t) = h_{\nu-1}(t) + h_{\nu-1}(t-a_k) \quad , \nu=1,..,n \ , \ t\in R^k$$

(8.23) mit $h_o(t) = (h_{o1},...,h_{ok})'$

und

$$h_{o1} = \begin{cases} 1 & t_1 = 0 \\ 0 & t_1 \neq 0 \end{cases} \qquad l=1,..,k \quad .$$

Die Komponenten T_1 haben jeweils $1+\sum_{i=1}^{n} a_{1i}$ mögliche Werte, $t_1=0,.., \sum_{i=1}^{n} a_{1i}$.

Diese kompliziertere k-dimensionale Verschiebe-Operation wird leichter verständlich, wenn man ein kleines künstliches Beispiel für den Spezialfall (8.21) betrachtet. Im Fall des unverbundenen Zweistichprobenproblems hat man:

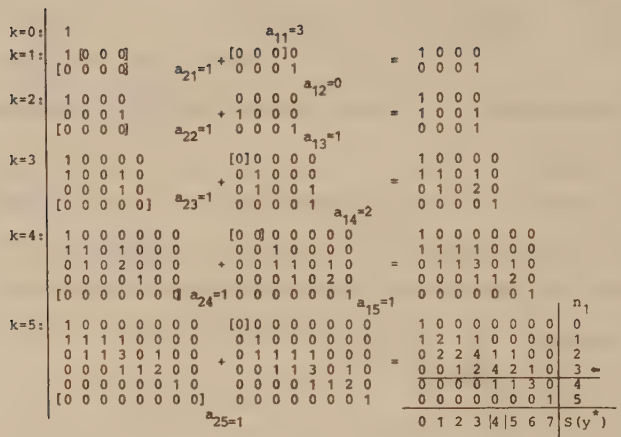
(8.24) $$A = \begin{pmatrix} y_1^* & y_2^* & \cdots & y_n^* \\ 1 & 1 & \cdots & 1 \end{pmatrix} \quad \text{und} \quad T = (T(y^*),N_1)' \quad .$$

Es müssen somit zweidimensionale Verschiebungen erfolgen.

Sei der transformierte 5-dimensionale Beobachtungsvektor

$$y^*=(3,0,1|2,1) \quad \text{mit } n_1=3 \text{ und } n_2=2 \ ; \ \sum_{i=1}^{5} y_i^* = 7 \quad .$$

Die sukzessive Ausführung der fünf 2-dimensionalen Verschiebungen verlangt jeweils Verschiebungen in der ersten Richtung um $a_{1i} = Y^*_i$ nach 'rechts' und simultan Verschiebungen um den konstanten Betrag von $a_{2i} = 1$ nach 'unten'. Die Rekursion beginnt wieder mit 1. Das Ergebnis ist in Tabelle 8.6 dargestellt.



Tab. 8.6: Beispiel: Zweidimensionaler Verschiebealgorithmus für das unverbundene Zweistichprobenproblem. Die absoluten Häufigkeiten für die Werte der Teststatistik $S(y^*)$ stehen in der unterstrichenen Zeile.

Die Häufigkeit, mit der für einen Stichprobenumfang von n_1 für die erste Stichprobe die verschiedenen Realisationen der Permutationsteststatistik (8.18) auftreten, liest man in der $n_1+1=4.$-ten Zeile der Matrix in der unteren rechten Ecke von Tabelle 8.6 ab. Für den beobachteten Wert $S(y^*) = 4$ ergibt sich eine Häufigkeit von 4 und damit eine Wahrscheinlichkeit von

$$P(S(y^*)=4 \mid n_1=3) = 1/\binom{n}{n_1} \cdot h((4,3)') = 4/10 \quad .$$

Der einseitige p-Wert ist dann $1/10 \cdot \sum_{t_1=4}^{7} h(\binom{t_1}{3}) = 7/10$.

Man erhält so für jede mögliche Anzahl n_1 von Elementen in der 1. Stichprobe $0<n_1<n$ die exakte bedingte Permutations-

verteilung von $S(y^*)$ durch Multiplikation der Häufigkeiten

$h((t,n_1)')$, $0 \le t \le \sum_{i=1}^{n} y_i^*$, mit $1/\binom{n}{n_1}$.

Benötigt man nur die Häufigkeit für ein festes n_1, so gibt es noch einige Vereinfachungen der Berechnung (Streitberg & Röhmel, 1986). Dann muß man nur die Zeilen $0 - n_1$ berechnen (angedeutet durch die waagerechten Striche in den Schritten 4 und 5), und es werden nur die Spalten für den Bereich

(8.25) $$t \le t_o := \sum_{j=n+1-n_1}^{n} y^*_{(j)}$$

benötigt. Im Beispiel ist das bei nach ansteigender Größe geordneten $Y^*_{(1)} \le .. \le Y^*_{(n)}$:

$t \le t_o = 1+2+3 = 6$.

Weitere Einsparungen sind bei Streitberg & Röhmel (1986, S. 14) beschrieben.

In Abhängigkeit von der Größe von $1+\sum_{i=1}^{n} y_i^*$ kann die Spaltenanzahl des Haufigkeitsfeldes erneut sehr groß ausfallen.

Ein ähnlicher Algorithmus stammt von Zimmermann (1983b). Für nicht zu große Gesamtstichprobenumfänge mit einer Anzahl verschiedener Werte von nicht mehr als 20 ist auch der Algorithmus von Stucky & Vollmar (1976), der keinen Verschiebe-Algorithmus benutzt, sehr geeignet (vgl. König, Giani & Haux, 1986).

Streitberg & Röhmel (1986) berichten auch über einen noch nicht ausgereiften sogenannten 'Shift-Partitions-Algorithmus', der keine zweidimensionalen Verschiebungen mehr benötigt.

In der englischsprachigen Fachliteratur ist für beide Zweistichprobenprobleme nur der Algorithmus von Pagano & Tritchler (1983) bzw. Tritchler (1984) für die zusätzliche Bestimmung eines Konfidenzintervalls bekannt. Dieser enthalt

an zentraler Stelle (Pagano & Tritchler, 1983, Theorem 1) ebenfalls eine Verschiebe-Operation, sodaß wiederum mit steigendem (Gesamt-) Stichprobenumfang der erforderliche Rechenaufwand nicht so schnell wächst wie für die vollständige explizite Ausführung aller unter der Nullhypothese zulässigen Permutationen. Der Algorithmus ist allerdings komplizierter als die zuvor dargestellten, da die Häufigkeiten, mit denen bestimmte Teststatistikwerte auftreten, nicht direkt bestimmt werden. Vielmehr wird zuerst (an endlich vielen Stützstellen) die sogenannte charakteristische Funktion der diskreten Wahrscheinlichkeitsverteilung der Teststatistik ermittelt und anschließend mit einer schnellen Fouriertransformation invertiert (vgl. Tritchler, 1984, Abschnitt 2). Die Teststatistiken sind im wesentlichen identisch mit (8.16) und (8.19).

Mit denselben Methoden ist es auch möglich, die gesamte exakte p-Wert Kurve (vgl. Abschnitt 6.2.4) bzw. die Teststatistik für jede Lagealternative (vgl. Tritchler, 1984) in Abhängigkeit vom Unterschied im Lageparameter beider Verteilungen zu bestimmen. Das ist gleichbedeutend mit der Möglichkeit, exakte $(1-\alpha)$-Konfidenzintervalle für die Differenz der Lageparameter zu berechnen. Für nicht zu kleine Stichprobenumfänge war das in Abschnitt 6.2.4 beschriebene Verfahren von Gabriel & Hall ja nicht mehr exakt, sondern nur mit einer Monte-Carlo Version (vgl. Abschnitt 8.3) ausführbar, da explizit alle zulässigen Permutationen der Beobachtungen benötigt wurden.

Andere Algorithmen zur effizienten exakten Konfidenzintervallbestimmung sind von John & Robinson (1983) und Heiler & Weichselberger (1969) sowie Zimmermann (1985) beschrieben

worden.

Für das (un-)verbundene k-Stichprobenproblem sind bisher keine ausgereiften Algorithmen publiziert worden. Eine Verallgemeinerung von Verschiebe-Algorithmen auf mehrdimensionale Verschiebe-Operationen (Streitberg & Röhmel, 1986; Pagano & Tritchler, 1983, Abschnitt 4) ist theoretisch möglich aber rechentechnisch sehr aufwendig und sehr speicherplatzintensiv.

Insgesamt scheint es in den letzten Jahren eine recht schnelle Entwicklung von effizienten, rechenzeitsparenden Algorithmen für kombinatorische Probleme zu geben (Victor, 1984). Streitberg & Röhmel (1986) schlagen deshalb für das unverbundene Zweistichprobenproblem einen sog. 'Polialgorithmus' vor, der je nach Gesamtstichprobenumfang, Balanziertheit der beiden Stichprobenumfänge und je nach Bereich zwischen kleinstem und größtem beobachtetem Wert den jeweils effizientesten Algorithmus einsetzt.

8.4 Spezielle Programme für Invarianz-Permutationstests

Nachdem im vorhergehenden Abschnitt die wesentlichen Bestandteile von (Invarianz-) Permutationstests in Form von Flußdiagrammen bzw. die grundlegenden Rekursionen von Verschiebe-Algorithmen dargestellt worden sind, sollen jetzt die in einer Sammlung von Einzelprogrammen für viele der zuvor behandelten Testprobleme erstellten FORTRAN-Algorithmen näher erläutert werden. Es wird jeweils ein Programm beispielhaft für einen Typ von Permutationstestprogrammen ausführlicher beschrieben. Die übrigen strukturähnlichen Programme für andere verwandte Testprobleme werden nur

zusammenfassend dargestellt. Auszuge aus einem Teil der Programm-Listings sind im Anhang B wiedergegeben. Alle Programme sind im FORTRAN 77 Standard geschrieben und auf einem VAX 11/780 Rechner mit einem VMS-Betriebssystem lauffahig. Bisher wurde kein großes Programmpaket eventuell mit einer eigenen Kontroll- und Steuersprache wie im SPSS, BMDP, SAS, usw. erstellt. Dazu sind viele algorithmische und numerische Probleme noch nicht endgültig gelöst. Es wurde jedoch darauf geachtet, daß bestimmte Unterprogramme z.B. für die systematische oder zufällige Erzeugung von zulässigen Permutationen der Beobachtungen oder für die Berechnung von Teststatistikwerten in multivariaten Testproblemen in möglichst vielen einzelnen Programmen Verwendung finden konnten. Oft ist die interaktive Auswahl zwischen verschiedenen Programm-Optionen über sog. Menü-Masken möglich. Die Spezifizierung von Ein- und Ausgabemedien erfolgt ebenfalls interaktiv. Felder sind in der Regel fest dimensioniert und mit entsprechenden Kommentarzeilen bezüglich der Anzahl von Faktorstufen (im allgemeinen maximal 10 je Faktor), Anzahl abhängiger Variablen (im allgemeinen maximal 10) und Stichprobenumfängen versehen. Wenn es möglich war, wurden bereits publizierte Unterprogramme verwendet.

8.4.1 Univariate Permutationstests

Pyhel (ohne Jahr) hat eine Sammlung von schnellen kombinatorischen Algorithmen zur Erzeugung aller in dieser Arbeit vorkommenden Gruppen von zulässigen Permutationen der Beobachtungen publiziert. Diese sind als Unterprogramme in

FORTRAN IV Standard angelegt und enthalten ihrerseits keine Aufrufe von weiteren Unterprogrammen z.B. zur Berechnung von Teststatistikwerten. Den umgekehrten Weg geht Berry (1982). Er ruft das kombinatorische Programm einmal auf. Alle Berechnungen der Teststatistikwerte sind in eigenen Routinen innerhalb dieses Unterprogramms realisiert.

Der Aufruf der einzelnen Programme ist möglichst einheitlich gehalten. Der Start und in einigen Unterprogrammen die Beendigung werden durch eine logische Variable START gesteuert, die vor dem ersten Aufruf einer Routine den Wert .TRUE. besitzen muß. Beim ersten Iterations-/Rekursionsschritt wird sie als .FALSE. definiert. Wenn ein Algorithmus n Iterationen enthält, wird erst beim (n+1)-ten Aufruf durch Setzen von START=.TRUE. abgebrochen.

Nachfolgend werden die verwendeten Algorithmen dargestellt. Zusätzlich sind die in dieser Arbeit beschriebenen Testprobleme, für deren Lösung sie Verwendung finden, angegeben. Generell gilt, daß für das zu einem univariaten Testproblem analoge multivariate Problem auf denselben Algorithmus zurückgegriffen werden kann. Es muß lediglich außerhalb des kombinatorischen Unterprogramms sichergestellt sein, daß jeweils die vektoriellen Beobachtungen als Ganze vertauscht werden (vgl. Abschnitt 4.3 und 4.5).

- KOMBIR: Kombinationen von N Elementen zu R Klassen ohne bzw. mit Berücksichtigung der Klassenanordnung: Verwendung in allen einfaktoriellen Plänen ohne Blocke/wiederholte Messungen. Der Algorithmus ist auch für den Split-Plot Plan verwendbar, da dort stets ganze Blocke bestehend aus den Werten über alle Split-Stufen zu vertauschen sind. Je nachdem ob die R Klassen gleiche bzw. unterschiedliche Besetzungen haben, werden alle $N!/n!^{R} \cdot R!$ bzw. $N!/\prod_{i=1}^{R} n_i!$

Kombinationen erzeugt. Dieser Algorithmus läßt verschiedene Permutationen nicht allein durch Inversionen auseinander hervorgehen, sondern benutzt rekursiv (R-1)-mal die Kombination von Elementen zu zwei Klassen. Bei den Trendtests und Tests für Kontraste ist stets auch für identische Stichprobenumfänge die Klassenanordnung zu berücksichtigen.

- KOMBI2: Kombinationen von N Elementen zu zwei Klassen vom Umfang n_1 und n_2: Verwendung in unverbundenen Zweistichprobenproblemen.

- BPERM: Erzeugung aller Matrizen mit R Zeilen und C Spalten, die durch Permutationen innerhalb der R Zeilen entstehen und bis auf Permutationen ganzer Spalten voneinander verschieden sind: Verwendung in allen einfaktoriellen Plänen mit Blockbildung/Meßwiederholungen, deren zugehörige Teststatistik invariant unter Spaltenpermutationen sind. Es werden alle $C!^{R-1}$ Permutationen innerhalb der Zeilen bei Festhalten der letzten Zeile ausgeführt. Für Trendtests kann der Algorithmus mit verändertem Aufruf - R muß durch R+1 ersetzt werden - ebenfalls verwendet werden. In diesem Fall ist die Teststatistik ja gerade nicht invariant unter Vertauschungen der Faktorstufen (vgl. Abschnitt 7.4). Der Algorithmus benötigt die Routine PERMNS unabhängig für die einzelnen Zeilen der Matrix. In dieser Routine werden nur durch Inversionen alle C! Vertauschungen von C Elementen in einer Zeile erreicht. In diesem schnellen Algorithmus ist stets die Angabe des Inversionsindex notwendig.

- PERM2: Vorzeichen von N Elementen: Verwendung in verbundenen Zweistichprobenproblemen.

- FPERM: Erzeugung aller D-Tupel von Matrizen $A_i, i=1,..,D$, mit je R Zeilen und C Spalten, die durch Zeilen- und Spalten-Permutationen innerhalb der Matrizen $A_1,...,A_{D-1}$ entstehen; A_D bleibt fest: Verwendung in zweifaktoriellen Blockplänen. Es werden also $(R!C!)^{D-1}$ Permutationen ausgeführt. Falls auch A_D permutiert werden soll, muß im Aufruf lediglich D durch D+1 ersetzt werden.

- RANPER: Zufällige Permutationen von N Elementen:

Verwendung in allen Versuchsplänen mit Ausnahme des verbundenen Zweistichprobenproblems. Der Algorithmus ist bereits in (8.11) wiedergegeben worden. Je Aufruf werden mit Hilfe der VAX-Bibliotheksfunktion RAN(I) insgesamt N auf (0,1) gleichverteilte (Pseudo-)Zufallszahlen mit folgendem sog. 'linearen Kongruenzgenerator' erzeugt (vgl. Kennedy & Gentle, 1980, Kap. 6.2; bzw. bezeichnet als ein gemischter Kongruenzgenerator bei Rubinstein, 1981, Kap. 2):

(8.26) $$U_{i+1} = ((a \cdot U_i + c) \bmod q)/q$$

mit $a=69069$, $c=1$ und $q=2^{32}$.

Für c=0 hat man den bekannten multiplikativen Kongruenzgenerator SUPER-DUPER von Marsaglia (1972), der nur in einem der stringenten Tests von Marsaglia (1985) nicht akzeptiert wird. Falls man einen anderen Zufallsgenerator wünscht, ist lediglich ein eigenes FUNCTION-Unterprogramm einzufügen. Als Startwert U_o wird eine große ungerade Zahl gewählt, z.B. durch Aufruf der VAX-Bibliotheksfunktion CPU_TIME, die die Zeit in Hundertstelsekunden ab Null Uhr im Rechner bestimmt. Will man RANPER in MC-Lösungen anstelle der exakten Lösungen mit BPERM und FPERM (sowie KOMBI2) nutzen, muß für eine der zulässigen Permutationen im MC-Test, wie in Abbildung 8.1 (Buchstabe E) dargestellt, RANPER in BPERM (R-1)-mal aufgerufen werden, um je eine unabhängige Permutation

innerhalb jeder der C-dimensionalen Zeilen zu liefern. In FPERM sind sogar (D-1)-mal unabhängig voneinander je eine Permutation des R-dimensionalen Zeilenvektors und des C-dimensionalen Spaltenvektors auszuführen fur eine zulässige Permutation in der MC-Lösung für den Test auf Wechselwirkungen im zweifaktoriellen Blockplan.

Als Analogon für KOMBIR im MC-Fall genügt jeweils ein Aufruf.

- RANSIG: Zufällige Vorzeichen von N Elementen: Verwendung im verbundenen Zweistichprobenproblem.

Je Aufruf wird ein Vektor der Länge N mit zufällig angeordneten Elementen von $\pm$ 1 erzeugt.

In Tabelle 8.7 sind die zur Erzeugung der systematischen Permutationen für den exakten Permutationstest sowie für NPERM = 20.000 zufällige Permutationen in MC-Fall benötigten CPU-Zeiten für einige Anzahlen von Faktorstufen(kombinationen) und Stichprobenumfänge für die verschiedenen Algorithmen angegeben.

Algorithmus	Parameter				Anzahl Permutationen	CPU-Zeit Sekunden
	k	n_1	n_2	n_3		
KOMBI2	2	3	3		20	0.10
	2	3	4		35	0.10
	2	3	6		84	0.13
	2	3	10		286	0.14
	2	5	5		252	0.12
	2	8	8		12870	1.25
	2	8	10		43758	3.60
	2	9	10		92378	8.23
	2	10	10		184756	16.45
Monte-Carlo	2	8	10		20000	17.68
	2	9	10		20000	18.31
	2	10	10		20000	19.79
KOMBIR	3	3	3	3	280	0.04
	3	3	4	5	27720	2.27
Monte-Carlo	3	3	4	5	20000	12.22
	k	n				
BPERM	3	4			216	0.03
	3	5			1296	0.09
	3	6			7776	0.55
	3	7			46656	3.24
	4	3			576	0.05
	4	4			13824	0.91
	4	5			331776	21.10
Monte-Carlo*	3	7			20000	5.72
	4	4			20000	2.98
	4	5			20000	3.83
PERM2	2	10			1024	0.06
	2	16			65536	3.94
Monte-Carlo*	2	10			20000	8.46
	2	16			20000	14.05
	r	c	n			
FPERM	2	2	5		256	0.02
	2	3	5		20736	1.56
	3	3	4		46656	3.53
Monte-Carlo*	2	3	5		20000	26.55
	3	3	4		20000	22.38

* Die Anzahl Permutationen selbst ist ein Vielfaches von 20000. Angegeben ist die Zeit für 20000 Permutationen der Daten des entsprechenden Versuchsplanes.

Tab. 8.7: Rechenzeiten zur Erzeugung aller Permutationen für verschiedene exakte Permutationstests und für die entsprechende Monte-Carlo Lösung mit 20000 zufälligen Permutationen der Beobachtungen im jeweiligen Versuchsplan. Rechner: VAX 11/780, Betriebssystem VMS.

Neben der möglichst rechenzeitsparenden Programmierung der Permutationsalgorithmen ist auch eine weitgehende Vermeidung numerischer Ungenauigkeiten (sog. Rundungsfehler, Kennedy & Gentle, 1970, Kap. 3.5 - 3.8) bei der Berechnung des jeweiligen Teststatistikwertes von großer Bedeutung. Hauptproblem ist die Abfrage auf Gleichheit zweier Gleitkomma (Real oder Double precision)-Zahlen für die Bestimmung des $p^{=}$-Anteiles des p-Wertes. Denn für Rechnungen im Real-Zahlen Modus gelten Axiome für reelle Zahlen wie das Kommutativ- und Assoziativgesetz nicht mehr. Insbesondere bedeutet das, daß selbst identische Werte bei Permutationen nicht mehr dieselbe Summe ergeben müssen (vgl. Kennedy & Gentle, 1970, S. 39).

Pyhel (ohne Jahr) hat dieses Problem dadurch gelöst, daß er als Beobachtungen nur ganze Zahlen (Integer-Zahlen) zuläßt und alle Divisionen in den Teststatistikformeln durch komplementäre Multiplikationen ersetzt wie in (8.10). Für mehrfaktorielle Pläne, in denen relativ viele Divisionen schon bei der Bestimmung der (maximal-)invariant transformierten Beobachtungen erforderlich sind, führt das dazu, daß schon für relativ kleine Gesamtstichprobenumfänge die Teststatistikwerte im Computer nicht mehr als Integer-Zahlen darstellbar sind. Eventuelle Laufabbrüche bei möglichst großer Rechengenauigkeit eines Programmes kann man folgendermaßen möglichst vermeiden. (i) Die Beobachtungen werden in ganze Zahlen umgewandelt. Dabei sollte für Beobachtungen mit Dezimalstellen überlegt werden, wieviele Dezimalstellen wirklich relevant sind. Seien das d Stück. Durch Multiplikation der Beobachtungen mit 10^d erhält man Zahlen, die keine signifikanten Nachkommastellen mehr enthalten und leicht rechnerintern in Integer-Zahlen umgewandelt werden können.

(ii) Weiterhin sollte man von allen Werten das Minimum über alle n Zahlen subtrahieren. Damit ist Null der kleinste Wert dieser 'neuen' Rohwerte y^*_{ij}. (iii) Dann bestimmt man den Teststatistikwert für die Anordnung der neuen Rohdaten y^*, die zum maximalen möglichen Teststatistikwert führt. Für die divisionsfrei dargestellte Teststatistik des unverbundenen k-Stichprobenproblems (vgl. 3.40):

$$S(Y^*) = \sum_{i=1}^{k} \prod_{\substack{l=1 \\ l \neq i}}^{k} n_l (nY^*_{i.} - n_i Y^*_{..})^2 \qquad (8.27)$$

ist das der Fall, wenn die n_1 kleinsten y^*_{ij} sich in der ersten Stichprobe, die n_2 nächstkleinsten in der zweiten Stichprobe usw. befinden. Für diese Konstellation von y^*_{ij} berechnet man $S(y^*)$ als Gleitkommazahl und prüft, ob der resultierende Wert kleiner als die größte darstellbare Integer-Zahl ausfällt. Dann kann $S(Y^*)$ im Ganzzahlmodus berechnet und der p-Wert problemlos bestimmt werden.
(iv) Wenn max $S(y^*)$ zu groß ist, bleibt nur die in Abschnitt 8.3.1 bereits angedeutete Möglichkeit, mögliche Rundungsfehler bei der Abfrage auf $S(\pi y^*) = S(y^*)$ zu berücksichtigen. Da alle Beobachtungen bereits in Zahlen ohne Dezimalstellen überführt sind, kann man, wenn man z.B. die Teststatistik (8.27) anstelle von (3.40) benutzt, alle 'permutierten' Teststatistikwerte, die sich um weniger als EPS=0.5 von $S(y^*)$ unterscheiden, zu den Elementen der Menge mit $S(\pi y^*) = S(y^*)$ zählen. (v) Falls die Komponenten von y^* so groß sind, daß auch im Real-Modus $S(Y^*)$ nicht mehr darstellbar ist, oder falls ein multivariates Problem vorliegt (vgl. Abschnitt 8.4.2), ist die maximal tolerable Ungenauigkeit EPS eventuell in Abhängigkeit von der minimal auftretenden Differenz in den vorliegenden Daten zu wählen bzw. in (iv) mit

um eine oder mehrere Zehnerpotenzen reduzierten Daten zu arbeiten. Die dadurch entstehende Ungenauigkeit sollte aber inhaltlich vertretbar sein. Standardmäßig ist EPS = 10^{-4} gesetzt. Eine Abschätzung des maximal möglichen relativen oder absoluten Fehlers mit den z.B. in Kennedy & Gentle (1970, Kap. 3.5 - 3.6) beschriebenen Methoden erscheint für die relativ komplexen Formeln in den Teststatistiken wenig sinnvoll.

Mit Ausnahme des (un-)verbundenen Zweistichprobenproblems und den Trendtests, in dem ein- und zweiseitige Alternativhypothesen wählbar sind, sind alle anderen Alternativen in den ein- oder multifaktoriellen Plänen einseitig. Nur 'große' Werte der Teststatistiken sind Anzeichen von Abweichungen von der Nullhypothese. Stets werden nur nichtrandomisierte p-Werte bestimmt. Da in den Ergebnisausdrucken $p^{>}$, $p^{=}$ und die Anzahl an ausgeführten Permutationen angegeben sind, kann anhand von (8.12) ein randomisierter p-Wert bestimmt werden.

Im (un-)verbundenen Zweistichprobenproblem sind die Algorithmen von Pagano & Tritchler (1983) und Streitberg & Röhmel (1986) (vgl. Abschnitt 8.3) zusätzlich zu den Permutationsalgorithmen wählbar sowie die p-Wert Approximation von John & Robinson (1983a, vgl. Abschnitt 8.1).

8.4.2 Multivariate Permutationstests

Wie im vorhergehenden Abschnitt bereits erwähnt, können die kombinatorischen bzw. Zufallspermutationen-Unterprogramme aus den analogen univariaten Programmen ohne Änderungen übernommen werden. Im multivariaten Fall werden ja stets

ganze d-dimensionale Vektoren anstelle univariater Komponenten auf dieselbe Weise wie bei den univariaten Testproblemen permutiert. Da stets zur parametrischen MANOVA analoge Teststatistiken bevorzugt werden und nicht etwa Summen von univariaten Teststatistiken (vgl. Abschnitt 4.2 (ii)), ist die Verwendung des Ganzzahl-Modus nicht möglich. Stets sind bei Spurkriterien eine (verallgemeinerte) Inverse, bei Determinanten-Kriterien Determinanten (oder alle Eigenwerte) und bei Maximalwurzelkriterien der maximale Eigenwert einer Matrix (vgl. (4.18)) zu berechnen. Diese Operationen sind i.a. nur mit Gleitkommazahlen ausführbar.

Zusätzlich zur Notwendigkeit schneller Permutationsalgorithmen treten im multivariaten Fall bei der Auswahl einer geeigneten Teststatistik Überlegungen hinsichtlich der Rechenzeit, die zur Bestimmumg eines Wertes der Teststatistik notwendig ist. Alle Teststatistiken, die die permutationsinvariante Summe aus Hypothesen- und Fehlermatrix enthalten, sind deshalb aus Rechenzeitgründen vorzuziehen. Wie bereits dargestellt, ist in diesem Sinne das Pillai'sche Spurkriterium optimal. Neben der einmaligen Berechnung einer verallgemeinerten Inversen mit einem sehr genauen von Boullion & Odell (1971; S. 74f) beschriebenen iterativen Algorithmus, ist lediglich neben der wiederholten Bestimmung von $\underline{S}_h(\pi Y)$ die Berechnung der Matrixspur des Produktes von $\underline{S}_h$ und $(\underline{S}_h + \underline{S}_e)$ auszuführen.

Für MRPP-Tests sind keine rechenzeitaufwendigen Operationen zur Teststatistikberechnung erforderlich. Allerdings müssen für jede neue zulässige Permutation der Daten z.B. im unverbundenen k-Stichprobenproblem $\sum_{i=1}^{k} \binom{n_i}{2}$ Abstände berechnet werden. Wirklich wenig aufwendig unter Rechenzeitgesichts-

punkten erscheint nur das Spurkriterium von Pillai. Seine guten Teststärke- und Robustheitseigenschaften im Normalverteilungsfall wurden bereits in Abschnitt 4.3 dargestellt. Wenn allerdings vermutet wird, daß die experimentellen Effekte auf eine oder wenige der analysierten abhängigen Variablen beschränkt bleiben, sollte alternativ oder zusätzlich zumindest das Maximalwurzelkriterium bestimmt werden. Die neu erstellten Computerprogramme sind so angelegt, daß eine oder mehrere Teststatistiken ausgewählt werden können. Auch bei einer MC-Lösung werden, falls mehrere Kriterien zu berechnen sind, für alle Teststatistiken dieselben zufällig ausgewählten Vektorpermutationen verwendet. Damit sind Unterschiede in den p-Werten nicht auf mangelnde Qualität des Zufallszahlengenerators zurückführbar. Bei den MRPP-Statistiken (Programme von Berry & Mielke, 1983a; Berry, Mielke & Wong, 1986; Iyer, Berry & Mielke, 1983) hat man zusätzlich stets die Wahl, auch die p-Wert Approximation ausführen zu lassen (vgl. Abschnitt 4.7). In Abb. 8.5 sind die wesentlichen Teile des Computer-Ausdruckes für die Beispieldaten von Wilkinson (1975) für ein 3-dimensionales unverbundenes 3-Stichprobenproblem angegeben.

MANOVA PERMUTATION TEST

K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY

*** MULTIVARIATE CR - 1F DESIGN ***

MRPP-MANOVA MIT WILKINSON-DATEN

* R A W - D A T A *

INPUT FILE : WILKINSON.DAT
DATA-FORMAT : (3F1.0)

		DEPENDENT VARIABLES 1	2	3
SAMPLE 1	1	5.000	6.000	4.000
	2	6.000	7.000	5.000
	3	6.000	7.000	3.000
	4	4.000	5.000	5.000
	5	5.000	4.000	2.000
SAMPLE MEAN :		5.200	5.800	3.800
SAMPLE 2	1	2.000	2.000	7.000
	2	3.000	3.000	5.000
	3	4.000	4.000	6.000
	4	3.000	2.000	4.000
	5	2.000	1.000	4.000
SAMPLE MEAN :		2.800	2.400	5.200
SAMPLE 3	1	4.000	3.000	4.000
	2	6.000	7.000	6.000
	3	3.000	3.000	6.000
	4	5.000	5.000	6.000
	5	5.000	5.000	4.000
SAMPLE MEAN :		4.600	4.600	4.600
OVERALL MEAN :		4.200	4.267	4.533

DIAGNOSTIC CHECK FOR ACCURACY OF GEN. INVERSE APLUS=(SH+SE)$^-$ OF A=SH+SE:

*** MAXDIFF (A X APLUS X A - A) : 0.44409E-15 ***

*** MONTE-CARLO SOLUTION USING 20000 PERMUTATIONS ***

RESULTS OF THE MULTIVARIATE PERMUTATION TEST :

DEPENDENT VARIABLES	PILLAI TRACE TEST-STAT	$P^>$	P=	P-VALUE
1 - 3	0.9472	285	1	0.0143
DEPENDENT VARIABLES	**ROY MAX ROOT TEST-STAT**	$P^>$	**P=**	**P-VALUE**
1 - 3	0.7303	330	1	0.0166
DEPENDENT VARIABLES	**MRPP TEST-STAT**	$P^<$	**P=**	**P-VALUE**
1 - 3	2.6804	276	1	0.0138

RESULTS OF THE UNIVARIATE PERMUTATION TESTS :

DEPENDENT VARIABLE	VALUE TEST-STAT	$P^>$	P=	P-VALUE
1	15.6000	102	66	0.0084
2	29.7333	200	19	0.0110
3	4.9333	3087	1012	0.2050
DEPENDENT VARIABLE	**URPP TEST-STAT**	$P^<$	**P=**	**P-VALUE**
1	1.1333	223	238	0.0231
2	1.6667	274	176	0.0225
3	1.2667	4225	4615	0.4420

MRPP-STATISTIC FOR EUCLIDEAN DISTANCES

P-VALUE APPROXIMATION USING A PEARSON III TYPE OF DISTRIBUTION FOR THE TEST-STATISTIC

S A M P L E	AVERAGE DISTANCE WITHIN SAMPLE
1	2.7309
2	2.5907
3	2.7196
OBSERVED DELTA	2.6004
EXPECTED DELTA	3.3779
VARIANCE OF DELTA	0.0407
SKEWNESS OF DELTA	-1.2593
VALUE OF TEST-STATISTIC	-3.1594
P-VALUE	0.0104

CPUTIME : 544.24 SEC

Abb. 8.5: Computer-Ausdruck (verkleinert) des Programms CR1FMANOVA für die Reanalyse der Daten von Wilkinson (1975).
Links: Rohdaten und Mittelwerte;
Rechts: Ergebnisse der multivariaten und univariaten Permutationstests sowie der p-Wert Approximation der MRPP-Prozedur

Es wurde eine Monte-Carlo Lösung mit 19999 zufällig erzeugten Permutationen sowie der identischen Permutation für das Pillai'sche Spurkriterium, das Maximalwurzelkriterium von Roy und die MRPP-Statistik gewählt. In diesem Beispiel sind alle drei p-Werte sehr ähnlich. Sie liegen zwischen .0138 und .0166. Sie stimmen auch recht gut mit dem approximativen p-Wert für die MRPP-Statistik von p=.0104 überein. Der p-Wert von p=.0176 für die F-Approximation des Pillai'schen Spurkriteriums (vgl. Kres, 1975, S. 139) ist bei einem F=3.32 und f_1=6, f_2=22 Freiheitsgraden sehr ähnlich mit dem Wert im Permutationstest. Nach der Tafel 2 in Kres (1975) müßte der maximale Eigenwert θ_{max}=.7303 etwas größer als der nicht vertafelte 99%-Schwellenwert für die Parameter m=0, s=2, n=4 sein. Für die exakte Lösung mit 15!/(5!3!·3)=126126 Permutationen ergibt sich für das Spurkriterum ein p=.01286 bei einer Rechenzeit von 367.37 CPU-Sekunden. Der p-Wert aus der Monte-Carlo Lösung liegt in diesem Fall außerhalb des 99%-Konfidenzintervalls aus Tab. B.1.

Mit einer vorgeschalteten Transformation der Beobachtungen durch eine orthogonale Kontrastmatrix sind MANOVA-Permutationsteststatistiken bzw. MRPP-Statistiken auch für die sog. Profilanalyse ohne weitere Änderungen anwendbar. Es muß lediglich zusätzlich sichergestellt sein, daß die Variablen eines Profils vergleichbar sind (vgl. Abschnitt 4.6). Im einfachsten Fall der Untersuchung von k-Gruppen von Profilen auf Parallelität werden die Daten wie im univariaten Split-Plot Plan mit c Split-Stufen eingelesen. Wenn man statt eines Tests auf Wechselwirkungen oder zusätzlich eine Profilanalyse interaktiv per Terminal auswählt, wird der Beobachtungsvektor mit der orthogonalen Kontrastmatrix, die

jeweils die Differenz benachbarter Profilelemente (vgl. (4.54) rechter Teil) erstellt, transformiert. Diese (c-1) dimensionalen Vektoren werden dann mit dem Permutationstest für das multivariate k-Stichprobenproblem analysiert. Programme für den Test auf verallgemeinerte Profilparallelität (vgl. 4.6 (iv)) und den Test auf multivariate Profilparallelität (vgl. 4.6 (v)) sind ebenfalls vorhanden. Statt einer der oben genannten MANOVA-Teststatistiken kann auch die MRPP-Teststatistik (4.67) von Berry & Mielke (1983a) ausgewählt werden.
In allen multivariaten Permutationstestprogrammen werden stets auch zusätzlich die p-Werte für alle einzelnen univariaten Testprobleme ermittelt.

8.4.3 Bestimmung der Teststärke

Beim Einsatz eines Permutationstests, der ja ein auf die tatsächlichen Beobachtungen eines Experimentes bedingter Test ist, kann, wie in Abschnitt 6.1 bereits dargestellt, streng genommen stets auch nur eine bedingte Teststärke etwa gegen Lagealternativen bestimmt werden. Auf die ebenfalls zuvor vorgeschlagene Möglichkeit, mit einem Simulationsexperiment unter Verwendung von Stichproben aus Residuenverteilungen apriori, d.h. vor Ausführung eines Experimentes, die Teststärke abzuschätzen, wird in diesem Abschnitt nicht weiter eingegangen. Die Flußdiagramme sowohl für die exakte konditionale wie für die naive Teststärkeberechnung sind in den Abbildungen 8.3 und 8.4 schon vorgestellt worden. Nur für sehr kleine Experimente wird eine exakte Bestimmung der

exakten konditionalen Teststärke möglich sein, da $|Q_n|^2$ Permutationen mit Berechnung des Teststatistikwertes notwendig sind. Das ist auch für univariate Zweistichprobenprobleme der Fall, für die mit den Methoden aus Abschnitt 6.2.4 die Teststärkekurve in Abhängigkeit von der Lagealternative bestimmt werden kann. Deshalb wird fast immer nur eine Schätzung der exakten konditionalen Teststärke mit MC-Methoden möglich sein. Bei der Festsetzung der Anzahl von zufällig ausgewählten Permutationen für eine einzelne p-Wert Bestimmung in Abschnitt B des Flußdiagramms aus Abb. 8.3 sowie der Anzahl von Replikationen des Experimentes mit zufällig permutierten Ausgangsdaten wird in den Computer-Programmen eine größere Genauigkeit - d.h. größere Permutationenzahl - für die einzelne p-Wert Bestimmung empfohlen (vgl. Abschnitt 6.3.1). Die naive Teststärkebestimmung erfordert $2|Q_n|$ Permutationen, sodaß dieselben Obergrenzen an Permutationen für eine exakte Berechnung gewählt werden können wie bei der ausschließlichen Bestimmung des p-Wertes. Bei Aufruf der in den beiden vorigen Abschnitten beschriebenen Programme erfolgt eine interaktive Abfrage, ob eine exakte oder naive Teststärkebestimmung gewünscht wird. Die exakte konditionale Teststärke wird nur exakt bestimmt, falls $|Q_n| \leq 10^3$ ist. Falls die Anzahl zulässiger Permutationen größer ist - und das wird in der überwiegenden Zahl der Testprobleme der Fall sein - , kann man die Anzahl Permutationen (ANZPERM) zur Bestimmung eines p-Wertes und die Anzahl (ANZPERMQ) der durch zufällige Permutation der Daten entstehenden Anordnungen der Beobachtungen wählen. Je nach Größe des Versuchsplanes sind die Kombinationen (10.000, 100-1.000) bzw. (1.000, 100) sinnvoll. Je nachdem, ob die (geschätzte) exakte Teststarke zu einem festen Niveau

oder dem empirischen p-Wert zu bestimmen ist, muß im Flußdiagramm der Abbildung 8.3 bei C α eventuell durch diesen p-Wert ersetzt werden. Bei der Ausgabe werden auch die Schatzwerte $\hat{\delta}$ für die Effektparameter mit ausgegeben.

Die naive Teststärke wird immer dann exakt bestimmt, wenn auch die p-Wert Berechnung exakt erfolgt (i.a. falls $|\Omega_n| \leq$ 20.000). Falls die naive Teststärke gegen ein festes α - Niveau berechnet werden soll, muß das $(1-\alpha)$-Quantil der exakten bzw. MC-Permutationsverteilung der Teststatistik ermittelt werden. Dazu sind alle Teststatistikwerte der Größe nach zu ordnen und der $[(1-\alpha)ANZPERM]+1$ -größte von ihnen muß gefunden werden. Dieser stellt das Quantil $s_{1-\alpha}$ dar, gegen das im Schritt C' in Abb. 8.4 zu vergleichen ist. Falls der $[(1-\alpha)ANZPERM]+2$-te Wert identisch mit $s_{1-\alpha}$ ist, wird solange in der geordneten Reihe der Teststatistikwerte vorwärts gesucht, bis ein gegenüber $s_{1-\alpha}$ größerer Wert auftritt; sei das $s_{1-\alpha^*}$ mit $\alpha^* < \alpha$. Dieser wird dann als kritischer Wert in C' verwendet und die Teststärke zu diesem i.a. wenig kleineren Niveau α^* ermittelt. Beide Niveaus, das nominelle Niveau α und das erreichbare Niveau α^* werden ausgegeben.

Falls die naive Teststärke auf den p-Wert bezogen werden soll, können diese letzten Programmschritte entfallen. Dann ist nach Durchlauf von A - G in Abb. 8.4 folgendermaßen zu verfahren. Aus den u.U. (maximal-) invariant transformierten Daten werden die Effekte, auf die sich der jeweilige Test bezieht, geschätzt und die (transformierten) Beobachtungen um diesen Effekt bereinigt (vgl. (6.32)). Die Effektschätzungen werden ausgegeben. Die resultierenden Residuen $\hat{u}$ werden dann wieder zulässigen Permutationen unterworfen, die

Effektschätzungen addiert und der Wert der Teststatistik bestimmt. Dieser wird mit dem ursprünglichen Wert der Teststatistik bzw. dem $(1-\alpha^*)$-Quantil verglichen. Die Anzahl von größeren Teststatistikwerten liefert dann den Schätzwert für die naive Teststärke.

Ein Beispiel für die native Schätzung der Teststärke zum beobachteten p-Wert im verbundenen k-Stichprobenproblem zeigt die Abb. 8.6.

```
******************************************************************
*                 PERMUTATION TEST PROCEDURE                      *
*      FOR THE DEPENDENT VARIABLES OF AN ANOVA-DESIGN             *
*                                                                 *
* K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY *
*                                                                 *
*        ***        RB - 1F TEST PROBLEM        ***               *
*                                                                 *
*                BEISPIEL LEHMANN-DATEN                           *
******************************************************************

                 ********************
                 * R A W  -  D A T A *
                 ********************

INPUT FILE   : LEHMANN.DAT
DATA-FORMAT  : (4F3.1)

                   -------------------------------------------
                   I            SPLIT  FACTOR                I
                   I      1         2         3         4    I
-------------------I-----------------------------------------I
I           I  1  I    23.100    22.700    22.500    22.600  I
I           I  2  I    57.600    53.200    53.700    53.100  I
I           I  3  I    10.500     9.700    10.800     8.300  I
I   DEP.  1 I  4  I    23.600    19.600    21.100    21.600  I
I           I  5  I    11.900    13.800    13.700    13.300  I
I           I  6  I    54.600    47.100    39.200    37.000  I
I           I  7  I    21.000    13.600    13.700    14.800  I
I           I  8  I    20.300    23.600    16.300    14.800  I
I-----------------------------------------------------------I------------I
I   MEAN    I     I    27.826    25.413    23.875    23.188  I    25.075  I
I-----------------------------------------------------------I------------I

           ESTIMATES OF EFFECT PARAMETER
           *****************************

 ---------------------------------------------------------
 I DEPENDENT  I            SPLIT  FACTOR                 I
 I VARIABLE   I      1         2         3         4     I
 -------------I------------------------------------------I
 I     1      I    2.750     0.337    -1.200    -1.888   I
 -------------I------------------------------------------I
```

```
******************************************************************
*                 PERMUTATION TEST PROCEDURE                      *
*      FOR THE DEPENDENT VARIABLES OF AN ANOVA-DESIGN             *
*                                                                 *
* K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY *
*                                                                 *
*        ***        RB - 1F TEST PROBLEM        ***               *
*                                                                 *
*                BEISPIEL LEHMANN-DATEN                           *
******************************************************************

     *** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***

        RESULTS OF THE UNIVARIATE PERMUTATION TESTS :
  ------------------------------------------------------------
  I DEPENDENT  I    VALUE     I        I        I            I
  I VARIABLE   I  TEST-STAT   I   P'   I   P=   I  P-VALUE   I
  I------------+--------------+--------+--------+------------+
  I     1      I   101.4325   I   282  I     1  I   0.0142   I
  I------------I--------------I--------I--------I------------I

     NAIVE POWER ESTIMATE    ( AT OBSERVED P-VALUE )   :
     -----------------------------------------------------
     I DEPENDENT  I    VALUE     I        I            I
     I VARIABLE   I  TEST-STAT   I   Q'   I  Q-VALUE   I
     I------------+--------------+--------+------------+
     I     1      I   101.4325   I  9335  I   0.4667   I
     I------------I--------------I--------I------------I

CPU-TIME : 161.62 SEC
```

Abb. 8.6: Computer-Ausdruck (verkleinert) des Programms RB1FPOWER für die naive Teststärke-Schätzung zum beobachteten α-Niveau für die Beispieldaten bei Lehmann (1975, S. 264).
Links: Rohdaten, Mittelwerte und Effektschätzung;
Rechts: Ergebnisse der Monte-Carlo Lösung des Permutationstests und der Teststärkeschätzung.

Reanalysiert wurden dazu die Beispieldaten in Lehmann (1975, S. 264) aus einem Experiment zur Effektivität von Hypnose. Bei 8 Versuchspersonen wurde in zufälliger Abfolge verlangt, jeweils eine der vier Emotionen Furcht, Gleichgültigkeit, Niedergeschlagenheit oder Gelassenheit (calmness) zu zeigen. Als abhangige Variable wurde das Hautpotential gemessen. Die Schatzungen der Faktorstufeneffekte sind ebenfalls im Ausdruck angegeben. Im Gegensatz zum Rangtest von Friedman mit einem p-Wert von mehr als 5% (vgl. Lehmann, 1975, S. 264) ergibt sich für die Monte-Carlo Lösung des Permutationstests mit 19999 zufallig ausgewählten Permutationen p=.0142. Die naive Schätzung der Teststärke nach (6.33) zum beobachteten α-Niveau von .0142 ergibt mit 0.47 einen recht niedrigen Wert.

Bei multivariaten Testproblemen ist wegen des großen Rechenaufwandes nur die naive Teststärkebestimmung programmiert. Fur die MRPP-Statistiken hat man zusätzlich die Moglichkeit, anstelle eines Permutationstests zur jeweiligen p-Wert Berechnung in Teil B der Abb. 8.3 den approximativen p-Wert anhand der ersten drei Momente der Permutationsverteilung der Teststatistik zu errechnen. Damit verkürzt sich die Rechenzeit zur Schätzung der 'exakten' Teststarke beträchtlich, und es sind großere Anzahlen als 100 - 1000 Permutationen der Ursprungsdaten moglich, um einen genaueren Schätzwert zu gewinnen. Für das (un-) verbundene Zweistichprobenproblem ist die Generierung der gesamten Teststärkekurve in Abhangigkeit vom Lageparameter mit Hilfe der Anker-Statistik von Gabriel & Hall (1983) in Abschnitt 8.3.1 beschrieben. Hier wird aber nur eine MC-Losung moglich sein.

8.4.4 Multiple Tests für die Komponenten multivariater Beobachtungen

Von den in den Abschnitten 5.2 und 5.3 dargestellten multiplen Permutationstests ist neben den bekannten Bonferroni- und Bonferroni-Holm-Verfahren, die keine besonderen neuen Algorithmen oder Permutationstestprogramme erfordern, die spezielle sequentielle Testprozedur von Interesse, die die Kovariation zwischen den einzelnen unabhängigen Variablen zur Testentscheidung nutzt und so die Teststärke verglichen mit der Bonferroni-Holm Prozedur im allgemeinen erhoht (vgl. Abb. 5.1). Im ersten Schritt dieses Verfahrens müssen die d 'univariaten' p-Werte bestimmt werden. Dieser Algorithmus ist analog zu dem bekannten Permutationstest-Algorithmus in Abb. 8.1. Es ist lediglich zu beachten, daß in B jeweils d univariate Teststatistikwerte für eine Vektorpermutation zu berechnen und in C jeweils mit dem jeweiligen ursprunglichen Teststatistikwert zu vergleichen sind. Entsprechend benötigt man einen d-elementigen 'Großer/Gleich'-Zähler und in F wird ein d-komponentiger Vektor von p-Werten berechnet. Wie bereits erläutert, konnen in E die 'univariaten' Permutationsalgorithmen nun angewendet auf d-dimensionale Vektoren eingesetzt werden. Damit erhoht sich lediglich der Rechenaufwand zur Bestimmung der Teststatistikwerte auf das d-fache, nicht jedoch die Zeit, die für die Generierung systematischer oder zufalliger Permutationen der Beobachtungen erforderlich ist. Weiterhin erhoht sich der erforderliche Speicherplatzaufwand, denn in C erfolgt nicht nur der Vergleich mit den d ursprunglichen Teststatistikwerten, sondern es mussen auch alle zu permutierten Beobachtungen gehörenden d-dimensionalen Vektoren

von Teststatistikwerten in einem zweidimensionalen Feld VERTEILU mit d·NPERM (= Anzahl Permutationen) Feldplätzen gespeichert werden.

Bevor die eigentliche multiple Testprozedur, wie sie in Abschnitt 5.3.1 anhand eines Beispiels veranschaulicht worden ist, ausgeführt werden kann, müssen noch drei weitere Felder angelegt werden. Das 2-dimensionale Feld VERTEILS (d, NPMAX) enthält die je abhängiger Variable nach steigender Große angeordneten Teststatistikwerte. Falls nicht große sondern kleine Teststatistikwerte gegen die Nullhypothese sprechen, ist entsprechend nach abnehmender Größe der Teststatistikwerte zu ordnen. Das zweidimensionale Feld PERMNUM (d,NPMAX) enthält je Dimension die zu den geordneten Teststatistikwerten je Dimension gehorenden Sequenznummern der ausgefuhrten Permutationen. Diese sind zur Bestimmung der Mengen M^1 (vgl. Abschnitt 5.3.1) erforderlich. Das Feld INDEX(d) schließlich enthält die Reihenfolge der univariaten p-Werte geordnet nach aufsteigender Große. In Abbildung 8.7 ist ein Flußdiagramm angegeben, das die wesentlichen Schritte der sequentiellen Testprozedur veranschaulicht. Der FORTRAN-Algorithmus selbst in dem Unterprogramm SEQMTP (Sequential multiple test procedure) ist im Anhang B3 aufgelistet.

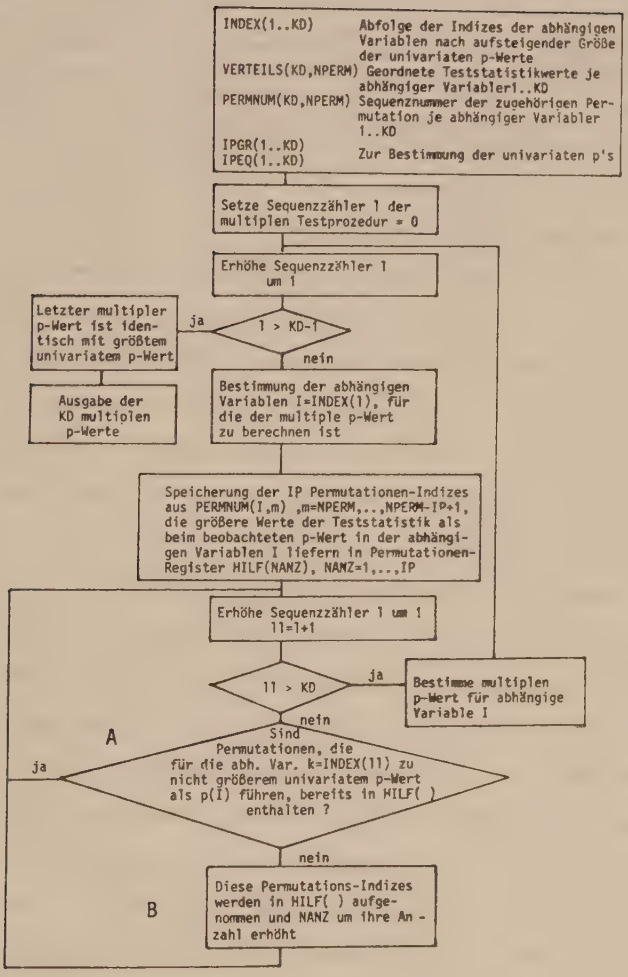


Abb. B.7: Vereinfachtes Flußdiagramm der sequentiellen multiplen Testprozedur für die abhängigen Variablen eines multivariaten Testproblems. Die Bezeichnungen entsprechen denen im Programm-Listing der Routine SEQMTP im Anhang B3.

In dem Flußdiagramm ist die Abfrage A selbst komplex. Zuerst ist zu prüfen, ob das empirische Signifikanzniveau für die abhangige Variable I = INDEX(1) (vgl. Abb. 8.7), die in der Sequenz der multiplen Testprozedur höher steht, auch in der ersten und den weiteren in der Sequenz nachfolgenden Dimension erreichbar ist, d.h. ob an dieser Stelle die empirische Verteilungsfunktion der nachgeordneten Dimensionen einen Sprung macht. Wenn nicht, muß bis zum nächsten Sprung in Richtung höherer Teststatistikwerte gezählt werden. Sind dann bis zum extremsten Teststatistikwert in diesen abhängigen Variablen noch Permutationsindizes übrig, muß geprüft werden, ob diese Indizes neu sind oder ob sie in dem Registerfeld HILF schon enthalten sind. Je nachdem, ob man eher Speicherplatz oder Rechenzeit in Abschnitt B investieren will, sind zwei alternative Vorgehensweisen möglich. In der speicherplatzintensiven Version des Algorithmus, die maximal $\sum_{i=1}^{d}(p_i^{>}+p_i^{=}) \leq d \cdot NPERM$ Registerplatze für die Permutationsindizes benötigt, kann man folgendermaßen verfahren. Man speichert alle Permutationsindizes je nach geordneter Dimension, ohne auf identische Indizes zu achten. Nach Durchlaufen aller nachgeordneten abhängigen Variablen wird erst dieses ganze Feld sortiert und alle identischen Indizes ausgeschieden. Die Anzahl verschiedener Permutationsindizes dividiert durch die Anzahl verwendeter Permutationen NPERM ergibt den multiplen p-Wert für die abhängige Variable I. Falls nur maximal NPERM Speicherplatze für die Permutationenindizes zur Verfügung stehen, muß stets sofort für jede in A ausgewählte Permutation durch Vergleich mit allen bisher in HILF gespeicherten Indizes entschieden werden, ob sie neu ist oder nicht. Insgesamt kann der Algorithmus beschleunigt werden, wenn man nur solange multiple p-Werte

ermittelt, wie diese nicht größer als ein vorgegebenes multiples Niveau α werden. Diese Programme sind so geschrieben, daß auch stets die Ergebnisse der Bonferroni-Holm Prozedur ausgegeben werden, die ja nur die univariaten p-Werte benötigt (vgl. Abschnitt 5.2). In Abbildung 8.8 ist der Computer-Ausdruck einer Reanalyse der Daten von Wilkinson (1975) aus einem 3-dimensionalen unverbundenen 3-Stichprobenproblem mit je 5 Beobachtungen angegeben.

Abb. 8.8: Computer-Ausdruck (verkleinert) des Programms CR1FMULTI der sequentiellen multiplen Testprozedur der drei abhängigen Variablen für die Wilkinson-Daten (vgl. Abb. 8.5).
Links: Exakte Lösung;
Rechts: Monte-Carlo Lösung.

Die Originaldaten sind Abbildung 8.5 zu entnehmen. Es wurden die exakte und eine Monte-Carlo Losung bestimmt und zu Demonstrationszwecken alle multiplen p-Werte ausgegeben, obwohl man bei Vorgabe eines multiplen α-Niveaus von z.B. 5% keinen multiplen p-Wert für die dritte abhangige Variable mehr angeben mußte. Vergleicht man die multiplen p-Werte der beiden ersten Variablen, die kleiner als das multiple α - Niveau ausgefallen sind, mit den entsprechenden p-Werten der sequentiellen, multiplen Testprozedur nach Holm, erkennt man die gleichmäßige Überlegenheit der neuen speziellen sequentiellen multiplen Testprozedur, wenn sie in diesem Fall auch nicht zu einer anderen Testentscheidung fuhrt. In diesem Beispiel sind alle univariaten p-Werte der Monte-Carlo Losung innerhalb der 99%-Konfidenzintervalle um die exakten p-Werte.

Für alle in Abschnitt 8.4.2 aufgeführten multivariaten Testprobleme gibt es eine Programmversion mit dieser sequentiellen Testprozedur.

Falls nicht Teilmengenhypothesen bzw. Hypothesen bezuglich der einzelnen abhängigen Variablen von Interesse sind, sondern die Hypothesenfamilien aller Linearkombinationen oder Kontraste bezuglich der abhangigen Variablen, kann das Flußdiagramm aus Abbildung 8.1 leicht zur Durchfuhrung einer simultanen Testprozedur (vgl. Abschnitt 5.3.2) modifiziert werden wie in Abbildung 8.9 dargestellt.

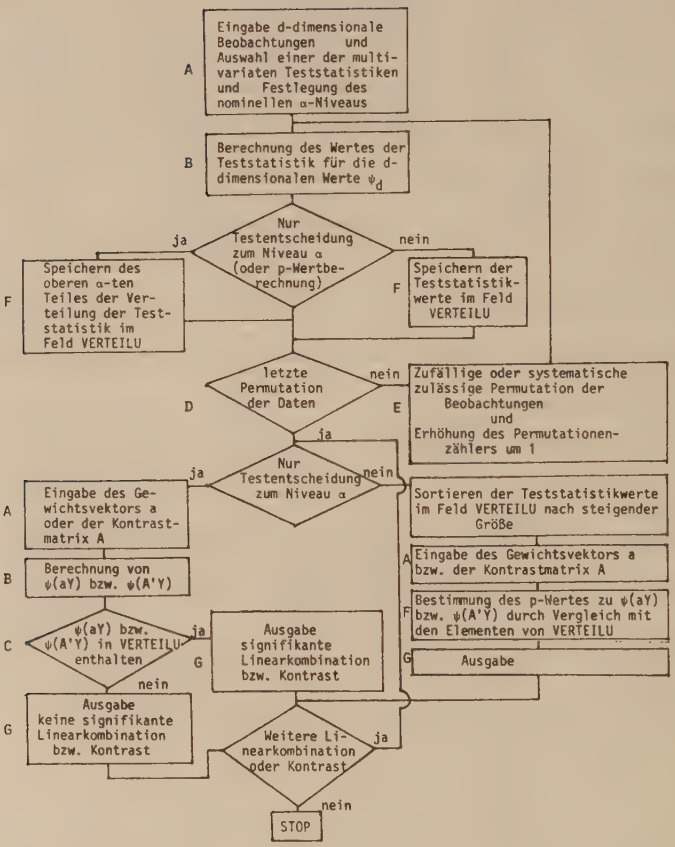


Abb. 8.9: Allgemeines Flußdiagramm für simultane Permutationstests aus Abschnitt 5.6.

Man führt nach Wahl einer multivariaten Teststatistik einen Permutationstest für alle d abhängigen Variablen aus. Es müssen im Gegensatz zur üblichen p-Wert Bestimmung in F zusätzlich alle Teststatistikwerte berechnet nach (4.18) aus B in einem Feld gespeichert werden. Je nachdem, ob man einen exakten oder MC-p-Wert für jede der betrachteten Hypothesen der Hypothesenfamilie aus (5.28) wünscht, oder nur eine Testentscheidung fällen will, muß die ganze empirische Verteilungsfunktion der Teststatistik durch Ordnen aller Werte in diesem Feld ermittelt werden bzw. lediglich ihr oberer α-ter Teil gespeichert werden mit Bestimmung des exakten $(1-\alpha)$-Quantils $\psi_{d,1-\alpha}$, falls dieses erreichbar ist. Dann müssen die Koeffizienten der Linearkombination(en) bzw. die Kontrastmatrizen interaktiv eingegeben oder eingelesen werden. Für jede dieser Teilhypothesen wird dann der Teststatistikwert $\psi^a(y)$ bzw. $\psi^A(y)$ nach (5.29) bestimmt. Durch Vergleich dieses Wertes mit den Werten der Verteilungsfunktion unter der globalen Nullhypothese bzw. mit dem $(1-\alpha)$-Quantil kann der p-Wert der Linearkombination/des Kontrastes bestimmt bzw. die Teilhypothese verworfen oder akzeptiert werden. Aufgrund der Eigenschaften einer simultanen Testprozedur können prinzipiell beliebig viele Teilhypothesen überprüft werden.

8.4.5 Multiple Tests für Faktorstufen

Wie für multiple Tests bezüglich der abhängigen Variablen in einem multivariaten Testproblem sind auch zum Vergleich von Faktorstufen für uni- und multivariate Probleme multiple sequentielle Testprozeduren besonders geeignet. Das von

Petrondas & Gabriel (1983) vorgestellte Verfahren für das univariate unverbundene k-Stichprobenproblem ist, wie bereits in Abschnitt 5.4 erlautert, nicht auf diesen Fall beschrankt. Falls die interessierenden Elementarhypothesen sich auf die Menge aller paarweisen Vergleiche von Faktorstufen beziehen, muß man neben der speziellen Wahl einer uni- oder multivariaten Teststatistik lediglich die Routine zur Erzeugung zulässiger Permutationen im speziellen Testproblem in dem von Petrondas & Gabriel (1981) publizierten FORTRAN-Programm auswechseln. In den eigenen modifizierten Routinen für das uni- und multivariate (un-) verbundene k-Stichprobenproblem sind jedoch mehrere Veränderungen eingeführt worden, die u.a. die benötigte Rechenzeit deutlich verringern. Diese zeigt auch der Computer-Ausdruck mit dem reanalysierten Beispiel von Petrondas & Gabriel (1983, Tab. 3) in Abbildung 8.10 bei Verwendung der 'F-Statistik' aus (5.32):

```
SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE
COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN
K.WILLMES & M.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY
***  UNIVARIATE CR - IF TEST PROBLEM  ***
PETRONDAS-GABRIEL DATEN
*** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***
F-STATISTIC
```

SUB-SET #	NUMBER OF PERMU-TATIONS	P-VALUES	LEVELS AT *) ALPHA=0.100 M-K	RYAN	DECISION *) AJ	A PERITZ	TREATMENTS INCLUDED IN SUBSET
			0.1000	0.1000		AJ	1 2 3 4 5 6 7
1	20000	0.0011					
			0.1000	0.0533			
2	20000	0.0159					1 2 3 4 5 6 0
3	20000	0.0033			A2		1 2 3 4 5 6 7
4	20000	0.0134			A2	A2	1 2 3 4 5 6 7
5	20000	0.0022					1 2 3 4 5 6 7
6	20000	0.0018			A2		1 2 3 4 5 6 7
7	20000	0.0224					1 2 3 4 5 6 7
			0.1000	0.0667			
8	20000	0.1097					1 2 3 4 5 6 0
9	20000	0.0158			A2		1 2 3 4 5 6 7
10	20000	0.0134			A2	A2	1 2 3 4 5 6 7
11	20000	0.0153			A2	A	1 2 3 4 5 6 7
12	20000	0.1425					1 2 3 4 5 6 0
13	20000	0.0134			A2	A2	1 2 3 4 5 6 7
14	20000	0.0050			A2		1 2 3 4 5 6 7
15	20000	0.0043			A2	A2	1 2 3 4 5 6 7
16	20000	0.0250			A2		1 2 3 4 5 6 7
17	20000	0.0209			A2	A2	1 2 3 4 5 6 7
18	20000	0.0087			A2		1 2 3 4 5 6 7
19	20000	0.0184			A2	A2	1 2 3 4 5 6 7
20	20000	0.0362			A2	A2	1 2 3 4 5 6 7
21	20000	0.0263			A2	A	1 2 3 4 5 6 7
22	20000	0.0763					0 1 2 3 4 5 6

```
SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE
COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN
K.WILLMES & M.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY
***  UNIVARIATE CR - IF TEST PROBLEM  ***
PETRONDAS-GABRIEL DATEN
*** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***
F-STATISTIC
```

SUB-SET #	NUMBER OF PERMU-TATIONS	P-VALUES	LEVELS AT *) ALPHA=0.100 M-K	RYAN	DECISION *) AJ	A PERITZ	TREATMENTS INCLUDED IN SUBSET
			0.1000	0.0500			1 2 3 4 5 6 7
23	0	------				A	1 2 3 4 5 6 7
24	0	------				A	1 2 3 4 5 6 7
25	0	------				A	1 2 3 4 5 6 7
26	0	------				A	1 2 3 4 5 6 7
27	5775	0.0050				A	1 2 3 4 5 6 7
28	5775	0.0039			A2	A	1 2 3 4 5 6 7
29	0	------				A	1 2 3 4 5 6 7
30	5775	0.0159			A2	A	1 2 3 4 5 6 7
31	0	------				A	1 2 3 4 5 6 7
32	0	------				A	1 2 3 4 5 6 7
33	5775	0.0030			A2	A	1 2 3 4 5 6 7
34	5775	0.0018			A2	A	1 2 3 4 5 6 7
35	5775	0.0039			A2	A	1 2 3 4 5 6 7
36	5775	0.0055			A2	A	1 2 3 4 5 6 7
37	5775	0.0112			A2	A	1 2 3 4 5 6 7
38	5775	0.0159			A2	A	1 2 3 4 5 6 7
39	5775	0.0023			A2	A	1 2 3 4 5 6 7
40	5775	0.0134			A2	A	1 2 3 4 5 6 7
41	5775	0.0332				A	0 1 2 3 4 5 6
42	5775	0.0254			A2	A	1 2 3 4 5 6 7
			0.1000	0.0333			
43	0	------				A	1 2 3 4 5 6 0
44	0	------				A	1 2 3 4 5 6 7
45	0	------				A	1 2 3 4 5 6 7
46	0	------				A	1 2 3 4 5 6 7
47	0	------				A	1 2 3 4 5 6 7
48	35	0.0286			A2	A	1 2 3 4 5 6 7
49	0	------				A	1 2 3 4 5 6 7
50	0	------				A	1 2 3 4 5 6 7
51	0	------				A	1 2 3 4 5 6 7
52	0	------				A	1 2 3 4 5 6 7
53	35	0.0286			A2	A	1 2 3 4 5 6 7
54	35	0.0331			A2	A	1 2 3 4 5 6 7
55	35	0.0331			A2	A	1 2 3 4 5 6 7
56	0	------				A	0 1 2 3 4 5 6
57	0	------				A	0 1 2 3 4 5 6

CPUTIME : 545.83 SEC

*) M-K : MENNAN-KEULS
RYAN : RYAN
AJ : ACCEPT
RJ : REJECT

Abb. 8.10: Computer-Ausdruck (verkleinert) des Programms GABRIEL-CR der multiplen sequentiellen Peritz-Prozedur im unverbundenen k-Stichprobenproblem bei Verwendung einer F-Statistik für die Beispieldaten von Petrondas & Gabriel (1983, Tab. 3; vgl. Abb. 8.12).

(i) Die aus dem Originalprogramm von Edgington (1980) stammenden Permutationsalgorithmen wurden durch KOMBIR und RANPER (vgl. Abschnitt 8.4.1) bzw. durch die entsprechenden Algorithmen für das verbundene k-Stichprobenproblem ersetzt.

(ii) In der Sequenz der zu testenden Teilmengenhypothesen beginnend mit k,k-1,... usw. Faktorstufen werden nur die nachgeordneten Permutationstests tatsachlich ausgeführt, die nicht schon aufgrund einer akzeptierten Teilmengenhypothese mit mehr Faktorstufen auch akzeptiert werden mussen. Genauer bedeutet das Folgendes: Nach Teil 1 von (5.40) müssen alle Hypothesen bzgl. Mengen von Faktorstufen, die in Hypothesen, die mit dem Verfahren von Newman & Keuls, d.h. zum Niveau α akzeptiert wurden, enthalten sind, ebenfalls in dem Peritz-Verfahren akzeptiert werden. In dem Beispiel von Petrondas & Gabriel sind das Hypothesen bezüglich folgender Teilmengen, da es bei Teilmengenhypothesen mit $|J(i)|=4$ zum erstenmal in der Sequenz nach Newman & Keuls zu akzeptierende Hypothesen J(8) und J(12) gibt: J(43)-J(48), J(23)-J(26) sind alle enthalten in J(8); J(45), J(47), J(48), J(50) - J(52), J(26), J(29), J(31), J(32) sind alle enthalten in J(12). Die mit p=.0763 'kritische' Hypothese J(22) (vgl. (5.40)) wird akzeptiert, da der Paarvergleich J(43) wegen J(8) zu akzeptieren war. Unter den Teilmengenhypothesen mit drei Faktorstufen gibt es weitere nach der Newman-Keuls Schranke (J(42)) oder nach dem Peritz-Verfahren (J(38), J(41)) zu akzeptierende Hypothesen. Bei den paarweisen Vergleichen bleiben zwei signifikante Ergebnisse für J(49) und J(53) übrig, wie auch bei der von Petrondas & Gabriel selbst verwendeten Bereichs-Statistik (vgl. (5.32)).

Für das angeführte Beispiel bedeutet die Nichtausführung 'überflüssiger' Tests eine beträchtliche Reduktion des Rechenaufwandes.

Gegenüber dem ursprünglichen Programm hat man noch die Möglichkeit, durch Ersetzen des Ryan-Niveaus $\alpha|J(i)|/k$ die Testprozedur weniger konservativ zu machen, ohne daß die Eigenschaft, ein multipler Test zum multiplen Niveau α zu sein, verletzt wird. Im Computer-Ausdruck sind diese mit

EINOT: $1-(1-\alpha)^{|J(i)|/k}$, $|J(i)|=k-1,..,2$;

WELSCH: Falls $|J(i)|=k-1$ wird $\alpha(k-1)/k$ ersetzt durch α; ansonsten wie bei Ryan;

EINOT +: Falls $|J(i)|=k-1$ wird $1-(1-\alpha)^{(k-1)/k}$ ersetzt durch α; ansonsten wie EINOT

bezeichnet. Das Beispiel zeigt auch, daß man die Teststärke verglichen mit der Bonferroni-Holm Prozedur i.a. erhöht. Keiner der beiden Paarvergleiche J(49) und J(53) wäre nach dieser Prozedur als signifikant abzusichern gewesen.

Wegen der großen Zahl zu testender Teilhypothesen ist für $10 \geq k > 6$ in den Programmen üblicherweise vorgesehen, nur alle Vergleiche von Paaren von Faktorstufen mit Interpretation der p-Werte nach der Bonferroni-Holm Prozedur vorzunehmen. Deren Verbesserung nach Hommel (1985) und Shaffer (1984, 1986) ist ebenfalls möglich.

Weiterhin wurde das in (5.41) angegebene step-up-Verfahren von Welsch (1977) programmiert. Dieses hat den Vorteil, daß selbst bei nur wenigen signifikanten Vergleichen von Paaren von Faktorstufen viele von den rechnerisch besonders aufwendigen Vergleichen von mehr als 2 Stufen nicht ausgeführt werden müssen, wenn man nicht an den exakten p-Werten, sondern nur an den Testentscheidungen interessiert ist. Die Abbildung 8.11 zeigt die Analyse der Beispieldaten von

Petrondas & Gabriel (vgl. Abb. 8.10):


```
SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE

COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN

K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY

***    UNIVARIATE CR - 1F TEST PROBLEM    ***

PETRONDAS-GABRIEL DATEN

*** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***

F-STATISTIC
```

SUB-SET #	NUMBER OF PERMU-TATIONS	P-VALUES	LEVELS AT *) ALPHA=0.100 M-K	RYAN	DECISION *) NJ A PERITZ	TREATMENTS INCLUDED IN SUBSET
			0.1000	0.1000	NJ	1 1 1 1 1 1
1	20000	0.0014				
			0.1000	0.0833		
2	20000	0.0158			NJ	1 1 1 1 1 0
3	20000	0.0034			NJ	1 1 1 1 0 1
4	20000	0.0004			NJ	1 1 1 0 1 1
5	20000	0.0021			NJ	1 1 0 1 1 1
6	20000	0.0224			NJ	0 1 1 1 1 1
			0.1000	0.0667		
8	20000	0.1098			A	1 1 1 1 0 0
9	20000	0.0057			NJ	1 1 1 0 1 0
10	20000	0.0179			NJ	1 1 1 0 0 1
11	20000	0.1412			NJ	1 1 0 1 1 0
12	20000	0.1412			NJ	1 1 0 1 0 1
13	20000	0.0015			NJ	1 1 0 0 1 1
14	20000	0.0238			NJ	1 0 1 1 1 0
15	20000	0.0039			NJ	1 0 1 1 0 1
16	20000	0.0306			NJ	1 0 1 0 1 1
17	20000	0.0009			NJ	1 0 0 1 1 1
18	20000	0.0179			NJ	0 1 1 1 0 1
19	20000	0.0106			NJ	0 1 1 0 1 1
20	20000	0.0037			NJ	0 1 0 1 1 1
21	20000	0.0731			NJ	0 0 1 1 1 1
22	20000	0.0778			A	0 0 1 1 1 1


```
SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE

COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN

K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY

***    UNIVARIATE CR - 1F TEST PROBLEM    ***

PETRONDAS-GABRIEL DATEN

*** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***

F-STATISTIC
```

SUB-SET #	NUMBER OF PERMU-TATIONS	P-VALUES	LEVELS AT *) ALPHA=0.100 M-K	RYAN	DECISION *) NJ A PERITZ	TREATMENTS INCLUDED IN SUBSET
			0.1000	0.0500		
23	5775	0.4044			A	1 1 1 0 0 0
24	5775	0.1380			A	1 1 0 1 0 0
25	5775	0.4691			A	1 1 1 1 0 0
26	5775	0.4635			A	1 1 1 0 1 0
27	5775	0.0859			A2	1 1 0 0 1 0
28	5775	0.0029			A2	1 1 0 0 0 1
29	5775	0.0386			A2	1 0 1 1 0 0
30	5775	0.1159			A2	1 0 1 0 1 0
31	5775	0.1236			A2	1 0 1 0 0 1
32	5775	0.3816			A2	1 0 0 1 1 0
33	5775	0.4018			A2	1 0 0 1 0 1
34	5775	0.4618			A2	1 0 0 0 1 1
35	5775	0.0159			A2	0 1 1 1 0 0
36	5775	0.0055			A2	0 1 1 0 1 0
37	5775	0.0133			A2	0 1 1 0 0 1
38	5775	0.0023			A2	0 1 0 1 1 0
39	5775	0.0023			A2	0 1 0 1 0 1
40	5775	0.0154			A2	0 1 0 0 1 1
41	5775	0.1272			A	0 0 1 1 1 0
42	5775	0.1254			A	0 0 1 1 0 1
			0.1000	0.0333		
43	35	0.2857			A	1 1 0 0 0 0
44	35	0.8571			A	1 0 1 0 0 0
45	35	0.2857			A	0 1 1 0 0 0
46	35	0.2857			A	1 0 0 1 0 0
47	35	0.0857			A	0 1 0 1 0 0
48	35	0.9143			A	1 0 0 0 1 0
49	35	0.4286			A2	0 1 0 0 1 0
50	35	0.4286			A2	1 0 0 0 0 1
51	35	0.1143			A	0 0 1 1 0 0
52	35	0.3143			A	0 0 1 0 1 0
53	35	0.3143			A2	0 0 1 0 0 1
54	35	0.0286			A	0 1 0 0 1 0
55	35	0.0571			A	0 0 1 0 0 1
56	35	0.0857			A	0 0 0 1 1 0
57	35	0.3714			A	0 0 0 1 0 1

CPUTIME : 549.45 SEC

*) M-K : NEWMAN-KEULS
RYAN : RYAN
A : ACCEPT
RJ : REJECT

Abb. B.11: Computer-Ausdruck (verkleinert) des Programms WELSCH_CR der multiplen sequentiellen Testprozedur von Welsch (1977) im unverbundenen k-Stichprobenproblem bei Verwendung einer F-Statistik für die Beispieldaten von Petrondas & Gabriel (1983; vgl. Abb. B.12)

Auch wenn im Beispiel die Programmversion mit Berechnung aller p-Werte gewählt wurde, sieht man leicht, daß in diesem Beispiel 10 der 20 Vergleiche von 3 Stufen (J(23)-J(26), J(32), J(35), J(37), J(38), J(41), J(42)), nur 2 der 15 Vergleiche von 4 Stufen (J(8), J(22)) und keiner der Vergleiche von 5 Stufen tatsächlich auszuführen wären, wenn man nur an einer Entscheidung bezüglich des multiplen Testniveaus interessiert ist. Im Unterschied zur Peritz-Prozedur ist bei den Paarvergleichen zusätzlich J(50) zu verwerfen sowie J(12), J(29) und J(31), obwohl der p-Wert von J(12) und J(31) größer als das Newman-Keuls Niveau von α=10% ausfällt. Das Welsch- und das Peritz-Verfahren sind also offensichtlich unterschiedlich teststark gegenüber unterschiedlichen Konstellationen von Faktorstufeneffekten. Untersuchungen zu dieser Problematik liegen auch für den parametrischen Fall bisher nicht vor.

In multivariaten multiplen Testproblemen kann man das modifizierte Programm von Petrondas & Gabriel auch nutzen, um in Abschnitt 5.6 beschriebene Permutationstests für Hypothesenfamilien bzgl. Komponenten und Faktorstufen auszuführen. Für jede nach dem Peritz-Verfahren verworfene d-dimensionale Hypothese über Teilmengen von Faktorstufen ist dann die in Abschnitt 8.4.4 beschriebene multiple Testprozedur für die Dimensionen zum Niveau $\alpha|J(i)|/k$, gegen das die multiplen p-Werte $p_{J(i)}^{[\nu]}$, $\nu=1,\ldots,d$, zu vergleichen sind, auszuführen. Routinemäßig ist nur bei nach dem Peritz-Verfahren signifikanten d-dimensionalen Paarvergleichen diese Analyse bzgl. der abhängigen Variablen vorgesehen. Die anderen in Abschnitt 5.6 beschriebenen Prozeduren für die allgemeineren Hypothesenfamilien sind nicht programmiert. Lediglich die Vergleiche gegen eine Kontrolle sind sowohl

univariat wie für den multivariaten Fall auch mit sequentieller Testprozedur für die Dimensionen als Programme vorhanden. Wie zuvor in der Peritz-Prozedur werden wieder nur alle die Teilhypothesen getestet, die nicht in bereits zum Niveau α akzeptierten Teilhypothesen enthalten sind. Falls man in dem Beispiel von Petrondas & Gabriel Stichprobe 6 als Kontrollbedingung auffaßt, ist das Ergebnis der in (5.45) angegebenen schrittweisen Testprozedur wie in Abbildung 8.12 dargestellt.

```
********************************************************************
*          SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE          *
*                                                                  *
*    COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN          *
*                                                                  *
* K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY *
*                                                                  *
*  *** UNIVARIATE MANY_ONE TEST PROBLEM FOR THE CR-1F DESIGN ***   *
*                                                                  *
*               MANY-ONE TEST MIT GABRIEL-DATEN                    *
********************************************************************

                      *******************
                      * R A W - D A T A *
                      *******************

INPUT FILE  : GABRIEL_CR.DAT
DATA-FORMAT : (F6.3)
```

SAMPLE 1	1	-2.885
	2	-0.298
	3	1.042
	4	2.141
SAMPLE MEAN		0.000
SAMPLE 2	1	0.105
	2	0.635
	3	5.008
	4	2.252
SAMPLE MEAN		2.000
SAMPLE 3	1	3.430
	2	4.208
	3	5.546
	4	1.456
SAMPLE MEAN		3.660
SAMPLE 4	1	5.708
	2	6.188
	3	0.181
	4	3.243
SAMPLE MEAN		3.830
SAMPLE 5	1	6.066
	2	5.019
	3	6.015
	4	6.910
SAMPLE MEAN		6.000
CONTROL	1	7.104
	2	8.572
	3	9.398
	4	4.926
CONTROL MEAN		7.500
OVERALL MEAN		3.832

```
********************************************************************
*          SEQUENTIAL MULTIPLE PERMUTATION TEST PROCEDURE          *
*                                                                  *
*    COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN          *
*                                                                  *
* K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY *
*                                                                  *
*  *** UNIVARIATE MANY_ONE TEST PROBLEM FOR THE CR-1F DESIGN ***   *
*                                                                  *
*               MANY-ONE TEST MIT GABRIEL-DATEN                    *
********************************************************************

     *** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***

     *** TEST FOR:  MUE 1,...,MUE i,...,MUE k-1  <  MUE k ***
```

SUB-SET NR	NUMBER OF PERMU-TATIONS	ALPHA= 0.050 P-VALUES	DECISION *) RJ	A	TREATMENTS INCLUDED IN SUBSET
1	20000	0.0005	RJ		1 1 1 1 1 1
2	20000	0.0003	RJ		1 1 1 1 0 1
3	20000	0.0007	RJ		1 1 1 0 1 1
4	20000	0.0010	RJ		1 1 0 1 1 1
5	20000	0.0005	RJ		1 0 1 1 1 1
6	20000	0.0040	RJ		0 1 1 1 1 1
7	20000	0.0003	RJ		1 1 1 0 0 1
8	20000	0.0006	RJ		1 1 0 1 0 1
9	20000	0.0005	RJ		1 0 1 1 0 1
10	20000	0.0036	RJ		0 1 1 1 0 1
11	20000	0.0010	RJ		1 1 0 0 1 1
12	20000	0.0003	RJ		1 0 1 0 1 1
13	20000	0.0032	RJ		0 1 1 0 1 1
14	20000	0.0009	RJ		1 0 0 1 1 1
15	20000	0.0058	RJ		0 1 0 1 1 1
16	20000	0.0349	RJ		0 0 1 1 1 1
17	5775	0.0036	RJ		1 1 0 0 0 1
18	5775	0.0009	RJ		1 0 1 0 0 1
19	5775	0.0076	RJ		0 1 1 0 0 1
20	5775	0.0033	RJ		1 0 0 1 0 1
21	5775	0.0192	RJ		0 1 0 1 0 1
22	5775	0.0687		A	0 0 1 1 0 1
23	5775	0.0029	RJ		1 0 0 0 1 1
24	5775	0.0099	RJ		0 1 0 0 1 1
25	5775	0.0324	RJ		0 0 1 0 1 1
26	5775	0.0464	RJ		0 0 0 1 1 1
27	35	0.0286	RJ		1 0 0 0 0 1
28	35	0.0571		A	0 1 0 0 0 1
29	0	------		A	0 0 1 0 0 1
30	0	------		A	0 0 0 1 0 1
31	35	0.2857		A	0 0 0 0 1 1

```
CPUTIME : 395.08 SEC           *)    A : ACCEPT
                                    RJ : REJECT
```

Abb. 8.12: Computer-Ausdruck (verkleinert) des Programms MANY_ONE_CR für eine multiple sequentielle Testprozedur für die Vergleiche gegen eine Kontrolle im unverbundenen k-Stichprobenproblem für die Beispieldaten von Petrondas & Gabriel (1983).
Links: Rohdaten und Mittelwerte;
Rechts: Ergebnisse der Testprozedur.

Bei einem α-Niveau von 5% ist nur der Vergleich der ersten Bedingung mit der 'Kontrollbedingung' bei Verwendung der Teststatistik $Q_k^{J(i)}(Y) = \min_{j \in J(i)} \{\bar{Y}_{j.}\} - \bar{Y}_k$ (vgl. (5.44)) signifikant. Wie man den p-Werten bei Petrondas & Gabriel (1983, Tab. 3) für die Paarvergleiche entnehmen kann, wären die Vergleiche 29 und 30 dort signifikant ausgefallen, wenn man α=10% betrachtet hätte. Im Fall des (einseitigen) Vergleichs gegen eine Kontrolle wären somit signifikante Unterschiede für die ersten 4 Faktorstufen gegenüber der sechsten Faktorstufe festzustellen gewesen.

Abschließend sollen dieselben Daten noch mit dem in den Abschnitten 5.1.2 und 5.5 beschriebenen Verfahren von Foutz analysiert werden. Diese Prozedur hält, wie bereits dargestellt, lediglich das globale Niveau α sowie die sog. marginalen α-Niveaus ein. Wie von Foutz empfohlen, werden möglicherweise auftretende Bindungen zwischen Teststatistikwerten per Zufall aufgehoben. Je nach Anzahl zulässiger Permutationen bei Gültigkeit der globalen Nullhypothese wird wieder eine Monte-Carlo Lösung mit 19999 zufällig ausgewählten Permutationen ausgeführt. Je nachdem wieviele Faktorstufen k und damit Paarvergleiche für die Hypothesenfamilie bestehend aus allen Paarvergleichen vorliegen, muß für ein vorgegebenes globales α-Niveau die Zahl NPERM=20000 eventuell leicht modifiziert werden, damit bei gleichen marginalen α_i für alle k(k-1)/2 Vergleiche von Paaren von Faktorstufen diese marginalen α_i tatsächlich erreichbare Fehlerniveaus sind und das globale α nicht überschritten wird. Beispielsweise ist für k=6 im unverbundenen k-Stichprobenproblem mit r=k(k-1)/2=15 und α=5% die Zahl M=0.05·20000=1000 auf 1005 zu erhöhen, damit M/r=67 ganzzahlig ausfällt und NPERM* möglichst nahe bei NPERM liegt (vgl. Abb. 8.13):

```
**********************************************************************
*                                                                    *
*                        FOUTZ PROCEDURE                             *
*                                                                    *
*        COMPARISON OF LEVELS OF FACTORS IN A (M)ANOVA-DESIGN        *
*                                                                    *
*  K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY RWTH AACHEN, GERMANY  *
*                                                                    *
*        ***     UNIVARIATE CR - 1F TEST PROBLEM       ***          *
*                                                                    *
*                                                                    *
*                        BEISPIEL GABRIEL-DATEN                      *
**********************************************************************

      *** MONTE-CARLO SOLUTION USING  20000 PERMUTATIONS ***

   OVERALL ALPHA-LEVEL                                : ALPHA =  0.0500

   MARGINAL ALPHA-LEVEL PER COMPARISON                : ALPHA*=  0.0033

   NUMBER OF PERMUTATIONS                             : NPERM =  20000

   ADJUSTED  NUMBER OF PERMUTATION, SO THAT
   THE ALPHA- AND ALPHA*-LEVELS ARE ATTAINABLE        : NPERM*=  20100

   OVERALL NUMBER OF TEST STATISTIC VALUES IN
   CRITICAL REGION                                    :     M =   1005

   NUMBER OF TEST STATISTIC VALUES IN CRITICAL
   REGION PER COMPARISON                              :     M*=     67

   NUMBER OF COMPARISONS                              :     R =     15


        COMPARISON
        OF SAMPLES          DECISION          P-VALUE
        ------------------------------------------------
          1 - 2          N.S.(GT ALPHA*)
          1 - 3          N.S.(GT ALPHA*)
          1 - 4          N.S.(GT ALPHA*)
          1 - 5          N.S.(GT ALPHA*)
          1 - 6             SIGN           0.0002
          2 - 3          N.S.(GT ALPHA*)
          2 - 4          N.S.(GT ALPHA*)
          2 - 5          N.S.(GT ALPHA*)
          2 - 6          N.S.(GT ALPHA*)
          3 - 4          N.S.(GT ALPHA*)
          3 - 5          N.S.(GT ALPHA*)
          3 - 6          N.S.(GT ALPHA*)
          4 - 5          N.S.(GT ALPHA*)
          4 - 6          N.S.(GT ALPHA*)
          5 - 6          N.S.(GT ALPHA*)

CPU-TIME :3188.27 SEC
```

Abb. 8.13: Computer-Ausdruck (verkleinert) des Programms FOUTZ_CR für die multiple Testprozedur von Foutz (1981) zum Vergleich aller Paare von k unabhängigen Stichproben für die Beispieldaten von Petrondas & Gabriel (1983).

Diese notwendige Veranderung in der Gesamtzahl zu erzeugender Permutationen wird vom Programm stets nach Spezifizierung des α-Niveaus und von NPERM vorgenommen. Für die Beispieldaten ist bei α=5% nur der Vergleich aus 1. und 6. Bedingung signifikant. Die weiteren zum multiplen α-Niveau aus der Welsch-Prozedur (vgl. Abb. 8.11) signifikanten Vergleiche erreichen das marginale Signifikanzniveau nicht.

Fur den exakten Permutationstest werden dagegen die marginalen Niveaus u.U. nicht immer exakt eingehalten sondern leicht verändert. Bei der Wahl des globalen Niveaus wird stets geprüft, ob pro Elementarhypothese bei der Bildung von M ein Wert von mindestens 1 resultiert. Ansonsten ist das gewählte globale nominelle Niveau mit dem vorliegenden Stichprobenumfang nicht erreichbar.

Obwohl das Verfahren von Foutz für sehr viele Testprobleme prinzipiell anwendbar ist, wie Foutz et al. (1985) zeigen, sind nur das uni- und multivariate (un-)verbundene k-Stichprobenproblem mit allen r=k(k-1)/2 paarweisen Vergleichen bzw. allen r=k-1 Vergleichen gegen eine Kontrolle programmiert.

Aastveit, H. (1983). On the effect of correlation between plots in randomized block experiments. Biometrical Journal, 25, 129-153.

Abelson, R.P., & Tukey, J.W. (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. The Annals of Mathematical Statistics 34, 1347-1369.

Addelman, S. (1970). Variability of treatments and experimental units in the design and analysis of experiments. Journal of the American Statistical Association, 65, 1095-1108.

Afflerbach, L. (1985). The pseudo-random number generators in Commodore and Apple microcomputers. Statistische Hefte, 26, 321-333.

Ahrens, H., & Läuter, J. (1981). Mehrdimensionale Varianzanalyse (2. Auflage). Berlin: Akademie-Verlag.

Albers, W., Bickel, P.J., & van Zwet, W.R. (1976). Asymptotic expansions for the power of distribution free tests in the one sample problem. The Annals of Statistics, 4, 108-156.

Alf, E.F., & Abrahams, N.M. (1972). Comment on component-randomization tests. Psychological Bulletin, 77, 223-224.

Anderson, T.W. (1958). An introduction to multivariate statistical analysis. New York: Wiley.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., & Tukey, J.W. (1972). Robust estimates of location. Princeton: Princeton University Press.

Anscombe, F.J. (1948). The validity of comparative experiments (with discussion). Journal of the Royal Statistical Society, Series A, 111, 181-211.

Arnold, H.J. (1964). Permutation support for multivariate techniques. Biometrika, 51, 65-70.

Azzalini, A., & Cox, D.R. (1984). Two new tests associated with analysis of variance. Journal of the Royal Statistical Society, Series B, 46, 335-343.

Bakan, D. (1966). The test of significance in psychological research. Psychological Bulletin, 66, 423-437.

Baker, F.B., & Collier, R.O. (1966a). Monte Carlo F-II: A computer program for analysis of variance F-tests by means of permutation. Educational and Psychological Measurement, 26, 169-173.

Baker, F.B., & Collier, R.O. (1966b). Some empirical results on variance ratios under permutation in the completely randomized design. Journal of the American Statistical Association, 61, 813-820.

Baker, F.B., & Collier, R.D. (1968). An empirical study into factors affecting the F-test under permutation for the randomized block design. Journal of the American Statistical Association, 63, 902-911.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., & Brunk, H.D. (1972). Statistical inference under order restrictions. London: Wiley.

Barnard, G.A. (1963). Discussion of Professor Barlett's paper. Journal of the Royal Statistical Society, Series B, 25, 294.

Barton, D.E., & David, F.N. (1962). Randomization bases for multivariate tests. I. The bivariate case. Randomness of n points in a plane. Bulletin of the International Institute of Statistics, 39, 455-467.

Bartholomew, D.J. (1961). Ordered tests in the analysis of variance. Biometrika, 48, 325-332.

Bartlett, M.S. (1939). A note on tests of significance in multivariate analysis. Proceedings of the Cambridge Philosophical Society, 35, 180-185.

Basu, D. (1980). Randomization analysis of experimental data: The Fisher randomization test. Journal of the American Statistical Association, 75, 575-589.

Bauer, P., & Hackl, P. (1985). The application of Hunter's inequality in simultaneous testing. Biometrical Journal, 27, 25-38.

Bauer, P., Hackl, P., Hommel, G., & Sonnemann, E. (1986). Multiple testing of pairs of one-sided hypotheses. Metrika, 33, 121-127.

Bay, K.S. (1971). Applications of multivariate analysis of variance to educational and psychological research: Hypothesis testing by MANOVA programs. Multivariate Behavioral Research, 6, 325-362.

Begun, J.M., & Gabriel, K.R. (1981). Closure of the Newman-Keuls multiple comparisons procedure. Journal of the American Statistical Association, 76, 241-245.

Bell, C.B., & Haller, H.S. (1969). Bivariate symmetry tests: parametric and nonparametric. The Annals of Mathematical Statistics, 40, 259-269.

Bell, C.B., & Smith, P.J. (1969). Some nonparametric tests for the multivariate goodness-of-fit, multisample, independence, and symmetry problems. In P.R. Krishnaiah (Ed.), Multivariate Analysis - II (pp. 3-23). New York: Academic Press.

Bellhouse, D.R. (1986). Randomization in the analysis of covariance. Biometrika, 73, 207-211.

Berenson, M.L. (1982a). A study of several useful tests for ordered alternatives in the randomized block design. Communications in Statistics - Simulation and Computation, B11, 563-581.

Berenson, M.L. (1982b). A comparison of several k sample tests for ordered alternatives in completely randomized designs. Psychometrika, 47, 265-280.

Berenson, M.L., Gross, S.T. (1981). Designing completely randomized experiments to detect ordered treatment effects: An empirical study. Communications in Statistics - Simulation and Computation, B10, 405-431 und Erratum, B10, 659-661.

Berry, K.J. (1982). Algorithm AS 179: Enumeration of all permutations of multi-sets with fixed repetition numbers. Applied Statistics, 31, 169-173.

Berry, K.J., & Mielke, P.W. (1983a). Computation of finite population parameters and approximate probability values for multi-response permutation procedures (MRPP). Communications in Statistics - Simulation and Computation, B12, 83-107.

Berry, K.J., & Mielke, P.W. (1983b). Moment approximations as an alternative to the F-test in analysis of variance. British Journal of Mathematical and Statistical Psychology, 36, 202-206.

Berry, K.J., & Mielke, P.W. (1984). Computation of exact probability values for multi-response permutation procedures (MRPP). Communications in Statistics - Simulation and Computation, B13, 417-432.

Berry, K.J., & Mielke, P.W. (1985). Computation of exact and approximate probability values for a matched pairs permutation test. Communications in Statistics - Simulation and Computation, B14, 229-248.

Berry, K.J., Mielke, P.W., & Wong, R.K.W. (1986). Approximate MRPP p-values obtained from four exact moments. Communications in Statistics - Simulation and Computation, B15, 581-589.

Besag, J., & Diggle, P.J. (1977). Simple Monte Carlo tests for spatial pattern. Applied Statistics, 26, 327-333.

Bhapkar, R.P., & Patterson, K.W. (1977). On some nonparametric tests for profile analysis of several multivariate samples. Journal of Multivariate Analysis, 7, 265-277.

Bhapkar, V.P., & Patterson, K.W. (1978). A Monte Carlo study of some multivariate statistics for profile analysis of several samples. Journal of Statistical Computation and Simulation, 6, 223-237.

Birnbaum, Z.W. (1970). Asymptotically distribution free statistics similar to Student's t. In L. Le Cam, J. Neyman, & E.L. Scott (Eds.), Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability (pp. 325 - 329). Berkeley: University of California Press.

Bock, R.D. (1966). Contribution of multivariate experimental designs to educational research. In R.B. Cattell (Ed.), Handbook of multivariate experimental psychology (pp. 820-840). Chicago: Rand McNally.

Bock, R.D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw Hill.

Bolles, R.C. (1962). The difference between statistical hypotheses and scientific hypotheses. Psychological Reports, 11, 639-645.

Bortz, J. (1977). Lehrbuch der Statistik für Sozialwissenschaftler. Berlin: Springer.

Boullion, T.L., & Odell, P.L. (1971). Generalized inverse matrices. New York: Wiley.

Box, G.E.P. (1953). Non-normality and tests on variances. Biometrika, 40, 318-335.

Box, G.E.P., & Andersen, S.L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption (with discussion). Journal of the Royal Statistical Society, Series B, 17, 1-34.

Box, G.E.P., & Watson, G.S. (1962). Robustness to non-normality of regression tests. Biometrika, 49, 93-106.

Box, J.F. (1978). R.A. Fisher: The life of a scientist. New York: Wiley.

Boyett, J.M., & Shuster, J.J. (1977). Nonparametric one-sided tests in multivariate analysis with medical applications. Journal of the American Statistical Association, 72, 665-668.

Bradley, J.V. (1968). Distribution-free statistical tests. Englewood Cliffs, New Jersey: Prentice-Hall.

Bradley, J.V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Bray, J.H., & Maxwell, S.E. (1982). Analyzing and interpreting significant MANOVA's. Review of Educational Research, 52, 340-367.

Bredenkamp, J. (1969). Über die Anwendung von Signifikanztests bei theorie-testenden Experimenten. Psychologische Beiträge, 11, 275-285.

Bredenkamp, J. (1970). Über Maße der praktischen Signifikanz. Zeitschrift für Psychologie, 177, 310-318.

Bredenkamp, J. (1972). Der Signifikanztest in der psychologischen Forschung. Frankfurt: Akademische Verlagsgesellschaft.

Bredenkamp, J. (1980). Theorie und Planung psychologischer Experimente. Darmstadt: Steinkopff.

Bredenkamp, J. (1982). Verfahren zur Ermittlung des Typs einer statistischen Wechselwirkung. Psychologische Beiträge, 24, 53-75.

Bredenkamp, J. (1983). Zum Problem der Validität des statistischen Schlusses. In J. Bredenkamp, & H. Feger (Hrsg.), Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Band 5: Hypothesenprüfung (S. 14-18). Göttingen: Hogrefe.

Bredenkamp, J. (1984). Anmerkungen und Korrekturen zu Hager & Westermann. Entscheidung über statistische und wissenschaftliche Hypothesen: Probleme bei mehrfachen Signifikanztests zur Prüfung einer wissenschaftlichen Hypothese. Zeitschrift für Sozialpsychologie, 15, 224-229.

Bredenkamp, J. & Erdfelder, E. (1985). Multivariate Varianzanalyse nach dem V-Kriterium. Psychologische Beiträge, 27, 127-154.

Brien, C.J. (1983). Analysis of variance tables based on experimental structure. Biometrics, 39, 53-59.

Brockwell, P.J., & Mielke, P.W. (1982). On non-normal invariance principles for multi-response permutation procedures. Australian Journal of Statistics, 24, 33-41.

Brockwell, P.J., & Mielke, P.W. (1984). Asymptotic distributions of matched-pairs permutation statistics based on distance measures. Australian Journal of Statistics, 26, 30-38.

Brown, B.M. & Maritz, J.D. (1982). Distribution-free methods in regression. Australian Journal of Statistics, 24, 318-331.

Brown, M.B., & Forsythe, A.B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. Biometrics, 30, 719-724.

Brunner, E., & Neumann, N. (1984). Rangtests für das verbundene Zwei-Stichprobenproblem mit fehlenden Meßwerten. Statistische Hefte, 25, 201-210.

Brunner, E., & Neumann, N. (1986). Two-sample rank tests in general models. Biometrical Journal, 28, 395-402.

Bunke, H., & Bunke, O. (1978). Randomization. Pro and contra. Mathematische Operationsforschung und Statistik, Series Statistics, 9, 607-623.

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campbell, G., & Skillings, J.H. (1985). Nonparametric stepwise multiple comparison procedures. Journal of the American Statistical Association, 80, 998-1003.

Carroll, R.M., & Nordholm, L.A. (1975). Sampling characteristics of Kelley's epsilon-square and Hays' omega-square. Educational and Psychological Measurement, 35, 541-554.

Cattell, R.B. (1966). Handbook of multivariate experimental psychology. Chicago: Rand McNally.

Chacko, V.J. (1963). Testing homogeneity against ordered alternatives. The Annals of Mathematical Statistics, 34, 945-956.

Chassan, J.B. (1960). 145 Note: On a test for order. Biometrics, 16, 119-121.

Chassan, J.B. (1962). 174 Note: An extension of a test for order. Biometrics, 18, 245-247.

Chung, J.H., & Fraser, D.A.S. (1958). Randomization tests for a multivariate two-sample problem. Journal of the American Statistical Association, 53, 729-735.

Cleroux, R. (1969). First and second moments of the randomization test in two-associate PBIB designs. Journal of the American Statistical Association, 64, 1424-1433.

Cochran, W.G., & Cox, G.M. (1957). Experimental designs (2nd ed.). New York: Wiley.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cole, J.W.L., & Grizzle, J.E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. Biometrics, 22, 810-829.

Collier, R.O., & Baker, F.B. (1963). The randomization distribution of F-ratios for the split-plot design - An empirical investigation. Biometrika, 50, 431-438.

Collier, R.O., & Baker, F.B. (1966). Some Monte Carlo results on the power of the F-test under permutation in the simple randomized block design. Biometrika, 53, 199-203.

Collier, R.O., Baker, F.B., & Mandeville, G.K. (1967). Tests of hypothesis in a repeated measures design from a permutation viewpoint. Psychometrika, 32, 15-24.

Collier, R.O., & Larson, R.C. (1969). Monte Carlo methods. Review of Educational Research, 39, 734-739.

Conover, W.J. (1971). Practical nonparametric statistics. New York: Wiley.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Boston: Houghton Mifflin.

Cox, D.F., & Kempthorne, O. (1963). Randomization tests for comparing survival curves. Biometrics, 19, 307-317.

Cox, D.R. (1956). A note on weighted randomization. The Annals of Mathematical Statistics, 17, 1144-1151.

Cox, D.R. (1958). Some problems connected with statistical inference. The Annals of Mathematical Statistics, 29, 357-372.

Cox, D.R. (1982). A remark on randomization in clinical trials. Utilitas Mathematica, A21, 245-252.

Cox, D.R., & Hinckley, D.V. (1974). Theoretical statistics. London: Chapman and Hall.

Cramer, E.M., & Nicewander, W.A. (1979). Some symmetric, invariant measures of multivariate association. Psychometrika, 44, 43-54.

Cran, G.W. (1980). Algorithm AS 149: Amalgamation of means in the case of simple ordering. Applied Statistics, 29, 209-211.

Dahme, B. (1984). Prüfung spezifischer Interventionshypothesen im Einzelfall mit Hilfe von Randomisierungstests. In D. Albert (Hrsg.), Bericht über den 34. Kongress der Deutschen Gesellschaft für Psychologie in Wien 1984, Bd. I (S. 147-149). Göttingen: Hogrefe.

Daniels, H.E. (1944). The relation between measures of correlation in the universe of sample permutations. Biometrika, 33, 129-135.

Davidson, M.L. (1972). Univariate versus multivariate tests in repeated-measures experiments. Psychological Bulletin, 77, 446-452.

Davis, A.W. (1980). On the effects of moderate multivariate nonnormality on Wilk's likelihood ratio criterion. Biometrika, 67, 419-427.

Davis, A.W. (1982). On the effects of moderate multivariate nonnormality on Roy's largest root test. Journal of the American Statistical Association, 77, 896-900.

De Groot, M.H. (1975). Probability und statistics. Reading Mass: Addison-Wesley.

Dempster, A.P., & Schatzoff, M. (1965). Expected significance level as a sensitivity index for test statistics. Journal of the American Statistical Association, 60, 420-436.

De Soete, G. (1982). A note on the use of critical levels and p-values for informative inference. Psychologische Beiträge, 24, 296-301.

Dolby, G.R. (1982). The role of statistics in the methodology of the life sciences. Biometrics, 38, 1069-1083.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. The Annals of Mathematical Statistics, 28, 181-187.

Dykstra, R.L. (1979). On dependent tests of significance in the multivariate analysis of variance. The Annals of Statistics, 7, 459-461.

Dykstra, R.L., & Robertson, T. (1983). On testing monotone tendencies. Journal of the American Statistical Association, 78, 342-350.

Easterling, R.G. (1975a). A comparison of some significance tests under experimental randomization. In J. N. Srivastava (Ed.), A survey of statistical design and linear models (p. 670). New York: North Holland/American Elsevier.

Easterling, R.G. (1975b). Randomization and statistical inference. Communications in Statistics - Theory and Methods, A4, 723-735.

Eden, T., & Yates, F. (1933). On the validity of Fisher's z test when applied to an actual example of non-normal data. Journal of Agricultural Science, 23, 6-17.

Edgington, E.S. (1964). Randomization tests. The Journal of Psychology, 57, 445-449.

Edgington, E.S. (1965). The assumption of homogeneity of variance for the t test and nonparametric tests. The Journal of Psychology, 59, 177-179.

Edgington, E.S. (1966). Statistical inference and nonrandom samples. Psychological Bulletin, 6, 485-487.

Edgington, E.S. (1967). Statistical inference from n = 1 experiments. The Journal of Psychology, 65, 195-199,

Edgington, E.S. (1969a). Statistical inference: The distribution free approach. New York: McGraw Hill.

Edgington, E.S. (1969b). Approximate randomization tests. The Journal of Psychology, 72, 143-149.

Edgington, E.S. (1970). Hypothesis testing without fixed levels of significance. The Journal of Psychology, 76, 109-115.

Edgington, E.S. (1971). Randomization tests with statistical control over concomitant variables. The Journal of Psychology, 79, 13-19.

Edgington, E.S. (1973). The random-sampling assumption in "comment on component-randomization tests". Psychological Bulletin, 80, 84-85.

Edgington, E.S. (1975a). Randomization tests for predicted trends. Canadian Psychological Review, 16, 49-53.

Edgington, E.S. (1975b). Randomization tests for one-subject operant experiments. The Journal of Psychology, 90, 57-68.

Edgington, E.S. (1980). Randomization tests. New York: Marcel Dekker.

Edgington, E.S., & Ezinga, G. (1978). Randomization tests and outlier scores. The Journal of Psychology, 99, 259-262.

Edgington, E.S. & Strain, A.R. (1973). Randomization tests: computer time requirements. The Journal of Psychology, 85, 89-95.

Einot, J., & Gabriel, K.R. (1975). A study of the powers of several methods of multiple comparisons. Journal of the American Statistical Association, 70, 574-583.

Erdfelder, E., & Bredenkamp, J. (1984). Kritik mehrfaktorieller Rangvarianzanalysen. Psychologische Beiträge, 26, 263-282.

Ferner, U. (1981). Simultane Hypothesenprüfungen. Bad Ischl: Biometrisches Seminar der Region österreich-Schweiz der Internationalen Biometrischen Gesellschaft.

Fienberg, S.E., & Hinckley, D.V. (1978). R.A. Fisher: An appreciation. Berlin: Springer.

Finch, P.D. (1979). Description and analogy in the practice of statistics. Biometrika, 66, 195-208.

Finch, P.D. (1982). Comments on paper by M. L. Thornett: The role of randomization in model based inference. Australian Journal of Statistics, 24, 146-147.

Finn, J.D. (1969). Multivariate analysis of repeated measures data. Multivariate Behavioral Research, 4, 391-413.

Fisher, R.A. (1925). Statistical methods for research workers. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1926). The arrangement of field experiments. Journal of the Ministry of Agriculture, 33, 503-513.

Fisher, R.A. (1935). The design of experiments. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1955). Statistical methods and scientific induction. Journal of the Royal Statistical Society, Series B, 17, 69-78.

Fisher, R.A., & Yates, F. (1963). Statistical tables for biological, agricultural and medical research (6th ed.). Edinburgh: Oliver & Boyd.

Forsythe, A.B., Engelman, L., Jennrich, R., & May, P.R.A. (1973). A stopping rule for variable selection in multiple regression. Journal of the American Statistical Association, 68, 75-77.

Forsythe, A.B., & Frey, H.L. (1970). Tests of significance from survival data. Computers and Biomedical Research, 3, 124-132.

Forsythe, A., & Hartigan, J.A. (1970). Efficiency of confidence intervals generated by repeated subsample calculations. Biometrika, 57, 629-639.

Foutz, R.V. (1980). A method for constructing exact tests from test statistics that have unknown null distributions. Journal of Statistical Computation and Simulation, 10, 187-193.

Foutz, R.V. (1981). Simultaneous tests for finite families of hypotheses. Communications in Statistics - Theory and Methods, A10, 1263-1275.

Foutz, R.V. (1984). Simultaneous randomization tests. Biometrical Journal, 26, 655-663.

Foutz, R.V., Jensen, D.R., & Anderson, G.W. (1985). Multiple comparisons in the randomization analysis of designed experiments with growth curve responses. Biometrics, 41, 29-37.

Franck, W.E. (1986). P-values for discrete test statistics. Biometrical Journal, 28, 403-406.

Gabriel, K.R. (1968). Simultaneous test procedures in multivariate analysis of variance. Biometrika, 55, 489-504.

Gabriel, K.R. (1969a). Simultaneous test procedures - Some theory of multiple comparisons. The Annals of Mathematical Statistics, 40, 224-250.

Gabriel, K.R. (1969b). A comparison of some methods of simultaneous inference in MANOVA. In P.R. Krishnaiah (Ed.), Multivariate analysis - II (pp. 67-86). New York: Academic Press.

Gabriel, K.R. (1979). Some statistical issues in weather experimentation. Communications in Statistics - Theory and Methods, A8, 975-1015.

Gabriel, K.R., & Hall, W.J. (1983). Rerandomization inference on regression and shift effects: Computationally feasible methods. Journal of the American Statistical Association, 78, 827-836.

Gabriel, K.R., & Hsu, Ch.-F. (1983). Evaluation of the power of rerandomization tests, with application to weather modification experiments. Journal of the American Statistical Association, 78, 766-775.

Gadenne, V. (1976). Die Gültigkeit psychologischer Untersuchungen. Stuttgart: Kohlhammer.

Gaebelein, J.W. , & Soderquist, D.R. (1978). The utility of within-subjects variables: Estimates of strength. Educational and Psychological Measurement, 38, 351-360.

Gammel, G., & Moosbrugger, H. (1982). Nichtparametrischer unbedingter Randomisierungstest für ein- und zwei-faktorielle Versuchspläne mit kleinen Stichprobengrößen. Psychologische Beiträge, 24, 253-276.

Garfield, S.L. (1978). Research problems in clinical diagnosis. Journal of Consulting and Clinical Psychology, 46, 596-607.

Geisser, S. (1980). Growth curve analysis. In P.R. Krishnaiah (Ed.), Handbook of Statistics 1: Analysis of variance (pp. 89-115). Amsterdam: North Holland.

Geisser, S., & Greenhouse, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. The Annals of Mathematical Statistics, 29, 885-891.

Gibbons, J.D., Olkin, I., & Sobel, M. (1977). Selecting and ordering populations: A new statistical methodology. New York: Wiley.

Gibbons, J.D., & Pratt, J.W. (1975). P-values: Interpretation and methodology. The American Statistician, 29, 20-25.

Gillies, D.A. (1971). A falsifying rule for probability statements. British Journal for the Philosophy of Science, 22, 231-261.

Giri, N. (1965). On the F-test in the intrablock analysis of a class of two associate PBIB designs. Journal of the American Statistical Association, 60, 285-293.

Gomez-Marin, O., & McHugh, R.B. (1984). Randomization modelling of the crossover experiment for clinical trials. Biometrical Journal, 26, 901-914.

Gordon, I. (1985). Error rates for orthogonal contrasts after a significant F-test. Australian Journal of Statistics, 27, 246-251.

Green, B.F. (1977). Practical interactive program for randomization tests of location. The American Statistician, 31, 37-39.

Green, B.F. (1981). Review of randomization tests by E.S. Edgington. Journal of the American Statistical Association, 76, 495.

Greenberg, B.G. (1957). Why randomize? Biometrics, 7, 309-322.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82, 1-20.

Gridgeman, N.T. (1959). The lady testing tea, and allied topics. Journal of the American Statistical Association, 54, 776-783.

Guenther, W.C. (1977). Power and sample size determination when the alternative hypotheses are given in quantiles. The American Statistician, 31, 117-118.

Guenther, W.C. (1981). Sample size formulas for normal theory t-tests. The American Statistician, 35, 243-244.

Guenther, W.C. (1982). Normal theory sample size formulas for some nonnormal distributions. Communications in Statistics - Simulation and Computation, B11, 727-732.

Guilford, J.P. (1956). Fundamental statistics in psychology and education. New York: McGraw Hill.

Hager, W. (1982). Zur statistischen Prüfung von Interaktionshypothesen. In G. Lüer (Hrsg.), Bericht über den 33. Kongreß der Deutschen Gesellschaft für Psychologie 1982, Bd. 1 (S. 197-201). Göttingen: Hogrefe.

Hager, W., & Möller, H. (1983). Tabellen und Verfahren zur Bestimmung der Teststärke und der Stichprobenumfänge in uni- und multivariaten Varianz- und Regressionsanalysen. Unveröffentlichtes Manuskript, Institut für Psychologie der Georg-August-Universität, Göttingen.

Hager, W., Lübbeke, B., & Hübner, R. (1983). Verletzung der Annahmen bei Zwei-Stichproben-Lokalisationstests: Eine Übersicht über empirische Resultate. Zeitschrift für experimentelle und angewandte Psychologie, 30, 347-386.

Hager, W., & Westermann, R. (1983a). Planung und Auswertung von Experimenten. In J. Bredenkamp & H. Feger (Hrsg.), Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Band 5: Hypothesenprüfung (S. 24-238). Göttingen: Hogrefe.

Hager, W., & Westermann, R. (1983b). Zur Wahl und Prüfung statistischer Hypothesen in psychologischen Untersuchungen. Zeitschrift für experimentelle und angewandte Psychologie, 30, 67-94.

Hager, W., & Westermann, R. (1983c). Entscheidung über statistische und wissenschaftliche Hypothesen: Probleme bei mehrfachen Signifikanztests zur Prüfung einer wissenschaftlichen Hypothese. Zeitschrift für Sozialpsychologie, 14, 106-117.

Harnatt, J. (1975). Der statistische Signifikanztest in kritischer Betrachtung. Psychologische Beiträge, 17, 595-612.

Hartigan, J.A. (1969). Using subsample values as typical values. Journal of the American Statistical Association, 64, 1303-1317.

Hartje, W., & Willmes, K. (1985). Activating right hemisphere functions by lateralized perceptual stimulation: a pilot study in patients with global aphasia. Unveröffentlichtes Manuskript, RWTH Aachen, Abteilung Neurologie, Aachen.

Hartley, H.O. (1977). Solution of statistical distribution problems by Monte Carlo methods. In K. Enslein, A. Ralston, & H.S. Wilf (Eds.), Statistical methods for digital computers (pp. 16-34). New York: Wiley.

Harville, D.A. (1975). Experimental randomization: Who needs it? American Statistician, 29, 27-31.

Hays, W.L. (1963). Statistics. London: Holt, Rinehart and Winston.

Hawel, W. (1982). Algorithm and FORTRAN program for the permutation test with dependent samples. In H. Caussinus, P. Ettinger, & J.R. Mathieu (Eds.), COMPSTAT 1982: Part II (Supplement): Short communications, summaries of posters (p. 133). Wien: Physica.

Heiler, S., & Weichselberger, K. (1969). Über den Permutationstest und ein daraus ableitbares Konfidenzintervall. Metrika, 14, 232-248.

Heller, R. (1981a). Ein FORTRAN-Programm zur Berechnung exakter p-Werte für Fisher-Pitman-Randomisierungstests. EDV in Medizin und Biologie, 12, 62-64.

Heller, R. (1981b). Pitman's one-way analysis of variance: exact p-values computed by a PL/1 program. EDV in Medizin und Biologie, 12, 81-82.

Henning, H.J., & Muthig, K. (1979). Grundlagen konstruktiver Versuchsplanung. München: Kösel.

Hicks, C.R. (1973). Fundamental concepts in the design of experiments (2nd edition). New York: Holt, Rinehard and Winston.

Hilgers, R. (1979). Ein asymptotisch verteilungsfreier Wechselwirkungstest in zweifaktoriellen vollständigen Zufallsplänen. Unveröffentlichte Dissertation, Universität Dortmund, Abteilung Statistik, Dortmund.

Hilgers, R. (1981). Anmerkungen zu den Tests von Conover und Iman. In: Symposium über nichtparametrische Methoden - Kurzfassungen - (S. 24-28), 09. - 11. November, München.

Hirotsu, C. (1978). Ordered alternatives for interaction effects. Biometrika, 65, 561-570.

Hirotsu, C. (1983). An approach to defining the pattern of interaction effects in a two-way layout. The Annals of the Institute of Statistics and Mathematics, A35, 77-90.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. The Annals of Mathematical Statistics, 23, 169-192.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.

Holschuh, N. (1978). Randomization and design: I., In S.E. Fienberg, & D.V. Hinckley (Eds.), R.A. Fisher: An appreciation (S. 35-45). Berlin: Springer.

Holtz, K.-L., & Eberle, G. (1978). Pädagogisch-therapeutische, experimentelle und statistische Signifikanz als Problem praktikabler Einzelfallanalyse. In DGVT Sonderheft II: Klinische Psychologie - Fortschritte in Diagnostik und Therapie. Kongreßbericht, Bd. 2, Hamburg.

Hommel, G. (1984). Multiple test procedures for arbitrary dependence structures. (Erscheint in Metrika).

Hommel, G. (1985). Multiple Vergleiche mittels Rangtests - alle Paarvergleiche. In G.Ch. Pflug (Hrsg.), Neuere Verfahren der nichtparametrischen Statistik (p. 28-48). Berlin: Springer.

Hope, A.C.A. (1968). A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society, Series B, 30, 582-598.

Horn, M. (1981). On the theory for Ryan's universal multiple comparison procedure with treatment of ties in the ranks of samples. Biometrical Journal, 33, 343-355.

Hotelling, H. (1931). The generalization of Student's ratio. The Annals of Mathematical Statistics, 2, 360-378.

Huber, P.J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35, 73-101.

Hubert, L.J. (1985). Combinatorial data analysis: Association and partial association. Psychometrika, 50, 449-467.

Hubert, L.J., & Baker, F. (1978). Evaluating the conformity of sociometric measurements. Psychometrika, 43, 31-41.

Hubert, L.J., & Golledge, R.G. (1983). Rater agreement for complex assessments. British Journal of Mathematical and Statistical Psychology, 36, 207-216.

Hubert, L., & Schultz, J. (1968). Quadratic assignment as a general data analysis strategy. British Journal of Mathematical and Statistical Psychology, 29, 190-241.

Huberty, C.J. (1972). Multivariate indices of strength of association. Multivariate Behavioral Research, 7, 523-526.

Hunter, D. (1976). An upper bound for the probability of a union. Journal of Applied Probability, 13, 597-603.

Hurlburt, R.T., & Spiegel, D.K. (1976). Dependence of F ratios sharing a common denominator mean square. The American Statistician, 30, 74-76.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. Psychometrika, 43, 161-175.

Huynh, H. (1981). Testing the identity of trends under the restriction of monotonicity in repeated measures designs. Psychometrika, 46, 295-305.

Huynh, H., & Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact F-distributions. Journal of the American Statistical Association, 65, 1582-1589.

Huynh, H., & Feldt, L.S. (1970). Estimations of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, 1, 69-82.

Ito, K. (1969). On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. In P.R. Krishnaiah (Ed.), Multivariate analysis - II (pp. 87-120). New York: Academic Press.

Ito, P.K. (1980). Robustness of ANOVA and MANOVA test procedures. In P.R. Krishnaiah (Ed.), Handbook of statistics 1: Analysis of variance (pp. 199-236). Amsterdam: North Holland.

Iyer, H.K., Berry, K.J., & Mielke, P.D. (1983). Computation of finite population parameters and approximate probability values for multi-response randomized block permutation procedures (MRBP). Communications in Statistics - Simulation and Computation, B12, 479-499.

Joe, G.W., & Anderson, H.E. (1969). Mean differences and association. Multivariate Behavioral Research, 4, 379-389.

Jöckel, K.-H. (1979). Sample size determination by simulation. EDV in Medizin in Biologie, 2, 59-61.

Jöckel, K.-H. (1980). Einige Aspekte zur Beurteilung statistischer Verfahren mittels Simulation. EDV in Medizin und Biologie, 11, 83-89.

Jöckel, K.-H. (1982). Eigenschaften und effektive Anwendung von Monte-Carlo-Tests. Dissertation, Universität Dortmund, Abteilung Statistik, Dortmund.

Jöckel, K.-H. (1986a). Monte Carlo tests. Statistical Software Newsletter, 12, 35-39.

Jöckel, K.-H. (1986b). Finite sample properties and asymptotic efficiency of Monte Carlo tests. The Annals of Statistics, 14, 336-347.

John, S. (1971). Some optimal multivariate tests. Biometrika, 58, 123-127.

John, R.D., & Robinson, J. (1983a). Significance levels and confidence intervals for permutation tests. Journal of Statistical Computation and Simulation, 16, 161-173.

John, R.D., & Robinson, J. (1983b). Edgeworth expansions for the power of permutation tests. The Annals of Statistics, 11, 625-631.

Johnson, N.L., & Kotz, S. (1970). Continuous univariate distributions. (Vol. 1 & 2). Boston: Houghton Mifflin.

Kadane, J.B. (1976). For what use are tests of hypotheses and tests of significance, introduction. Communications in Statistics - Theory and Methods, A5, 735-736.

Kanji, G.K. (1975). Robustness of power in the analysis of variance. Journal of Statistical Computation and Simulation, 4, 19-30.

Kempthorne, O. (1952). The design and analysis of experiments. New York: Wiley.

Kempthorne, O. (1955). The randomization theory of experimental inference. Journal of the American Statistical Association, 50, 946-967.

Kempthorne, O. (1966a). Some aspects of experimental inference. Journal of the American Statistical Association, 61, 11-34.

Kempthorne, O. (1966b). Multivariate responses in comparative experiments. In P.R. Krishnaiah (Ed.), Multivariate Analysis (pp. 521-540). New York: Academic Press.

Kempthorne, O. (1971). Probability, statistics and the knowledge business. In V.P. Godambe, & D.A. Sprott (Eds.), Foundations of Statistical Inference (p. 470-499). Toronto: Holt, Rinehart and Winston of Canada.

Kempthorne, O. (1972). Theories of inference and data analysis. In T.A. Bancroft (Ed.), Statistical papers in honor of G.W. Snedecor (p. 167-191). Ames: Iowa State University Press.

Kempthorne, O. (1975). Inference from experiments and randomization. In Srivastava, J.N. (Ed.), A survey of statistical design and linear models (pp. 303-331). New York: North Holland/American Elsevier.

Kempthorne, O. (1976). Of what use are tests of significance and tests of hypothesis. Communications in Statistics - Theory and Methods. A5, 763-777.

Kempthorne, O. (1977). Why randomize? Journal of Statistical Planning and Inference, 1, 1-25.

Kempthorne, O. (1980). Comment on D. Basu: Randomization analysis of experimental data: The Fisher randomization test. Journal of the American Statistical Association, 75, 584-587.

Kempthorne, O. (1982). Review of Randomization tests by E.S. Edgington. Biometrics, 38, 864-867.

Kempthorne, O. (1983). A review of R.A. Fisher: An appreciation. Journal of the American Statistical Association, 78, 482-490.

Kempthorne, O., & Barclay, W.D. (1953). The partition of error in randomized blocks. Journal of the American Statistical Association, 48, 610-614.

Kempthorne, O., & Doerfler, T.E. (1969). The behavior of some significance tests under experimental randomization. Biometrika, 56, 231-248.

Kempthorne, O., Zyskind, G., Addelman, S., Throckmorton, T.N., & White, R.F. (1961). Analysis of variance procedures. Aeronautical Research Laboratory Technical Report ARL 149. Wright-Patterson Air Force base, Ohio.

Kendall, M.G., & Stuart, A. (1968). The advanced theory of statistics, Vol. 3 (2nd ed.). London: Griffin.

Kendall, M.G., & Stuart, A. (1981). The advanced theory of statistics. Vol. 2: Inference and relationship (4th ed.). London: Griffin.

Kennedy, W.J., & Gentle, J.E. (1980). Statistical computing. New York: Marcel Dekker.

Kepner, J.L., & Robinson, D.H. (1984). A distribution-free rank test for ordered alternatives in randomized complete block designs. Journal of the American Statistical Association, 79, 212-217.

Keppel, G. (1973). Design and Analysis: A researcher's handbook. Englewood Cliffs: Prentice-Hall.

Kerber, A. (1971). Representations of permutation groups I. Lecture Notes in Mathematics, Vol. 240. Berlin: Springer.

Kimball, A.W. (1951). On dependent tests of significance in the analysis of variance. The Annals of Mathematical Statistics, 22, 600-602.

Kirk, R.E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, Cal.: Brooks/Cole.

Klauber, M.R. (1971). Two-sample randomization tests for space-time clustering. Biometrics, 27, 129-142.

Klauber, M.R. (1975). Space-time clustering tests for more than two samples. Biometrics, 31, 719-726.

Knuth, D.E. (1973). The art of computer programming, Vol. 3: Sorting and searching. Reading, Mass: Addison-Wesley.

Koch, G.G. (1969). Some aspects of the statistical analysis of 'split plot' experiments in completely randomized layouts. Journal of the American Statistical Association, 64, 485-505.

Koch, G.G. (1970). The use of non-parametric methods in the statistical analysis of a complex split plot experiment. Biometrics, 26, 105-128.

Koch, G.G., Amara, I.A., Stokes, M.E., & Gillings, D.B. (1980). Some views on parametric and non-parametric analysis for repeated measurements and selected bibliography. International Statistical Review, 48, 249-265.

König, J., Giani, G., & Haux, R. (1986). Some algorithms for the distributions of nonparametric two-sample tests - considerations on complexity and empirical investigations. Part 1: Permutation tests. Statistical Software Newsletter, 12, 20-27.

Koop, J.C. (1979). On statistical inference in sample surveys and the underlying role of randomization. Annals of the Institute of Statistical Mathematics, 31, 253-269.

Krafft, O. (1978). Lineare statistische Modelle und optimale Versuchspläne. Göttingen, Vandenhoeck & Ruprecht.

Krauth, J. (1982). Verteilungsfreie Homogenitätstests bei abhängigen Stichproben. Psychologische Beiträge, 29, 601-619.

Kres, H. (1975). Statistische Tafeln zur multivariaten Analysis. Berlin: Springer.

Krishnaiah, P.R. (1969). Multivariate analysis - II. New York: Academic Press.

Krishnaiah, P.R. (1980). Handbook of statistics 1: Analysis of variance. Amsterdam: North Holland.

Krishnaiah, P.R., Mudholkar, G.S., & Subbaiah, P. (1980). Simultaneous test procedures for mean vectors and covariance matrices. In P.R. Krishnaiah (Ed.), Handbook of statistics 1: Analysis of variance (pp. 631-671). Amsterdam: North Holland.

Kruskal, W. (1980). The significance of Fisher: A review of R.A. Fisher: The life of a scientist. Journal of the American Statistical Association, 75, 1019-1030.

Kshirsagar, A.M. (1972). Multivariate analysis. New York: Marcel Dekker.

Kubinger, K.D. (1986). A note on non-parametric tests for the interaction in two-way layouts. Biometrical Journal, 28, 67-72.

Kulatunga, D.D.S., & Sasabuchi, S. (1984). A test of simultaneous homogeneity against ordered alternatives in multifactorial design. Communications in Statistics - Theory and Methods, A13, 3173-3183.

Läuter, J. (1978). Sample size requirements for the T^2-test of MANOVA (tables for one-way classification). Biometrical Journal, 20, 389-406.

Lambert, D. (1982). Qualitative robustness of tests. Journal of the American Statistical Association, 77, 352-357.

Lambert, D. (1985). Robust two-sample permutation tests. The Annals of Statistics, 13, 606-625.

Lambert, D., & Hall, W.J. (1982). Asymptotic lognormality of p-values. The Annals of Statistics, 10, 44-64.

Lambert, D., & Hall, W.J. (1983). Asymptotic lognormality of p-values. The Annals of Statistics, 11, 348.

Lane, D.M., & Dunlap, W.P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. British Journal of Mathematical and Statistical Psychology, 31, 107-112.

Lee, H., & Fung, K.Y. (1985). Behavior of trimmed F and sine-wave F statistics in one-way ANOVA. Sankhya: The Indian Journal of Statistics, B47, 186-201.

Lehmann, E.L. (1959). Testing statistical hypotheses. New York: Wiley.

Lehmann, E.L. (1975). Nonparametrics: Statistical methods based on ranks. San Francisco: Holden-Day.

Lehmann, E.L., & Stein, C. (1949). On the theory of some non-parametric hypotheses. The Annals of Mathematical Statistics, 20, 28-45.

Lemmer, H.H. (1980). Some empirical results on the two-way analysis of variance by ranks. Communications in Statistics - Theory and Methods, A9, 1426-1438.

Lemmer, H.H., & Stoker, D.J. (1967). A distribution-free analysis of variance for the two-way classification. South African Statistical Journal, 1, 67-74.

Levin, J.R., Marascuilo, L.A., & Hubert, L.J. (1978). N = Nonparametric randomization tests. In T.R. Kratochwill (Ed.), Single subject research: strategies for evaluating change (pp. 167-196). New York: Academic Press.

Lindley, D.V., & Novick, M.R. (1981). The role of exchangeability in inference. The Annals of Statistics, 9, 45-58.

Lienert, G.A. (1973). Verteilungsfreie Methoden in der Biostatistik (Bd. I). Meisenheim am Glan: O. Hain.

Lorenzen, T.J. (1984). Randomization and blocking in the design of experiments. Communications in Statistics - Theory and Methods, A13, 2601-2623.

Lykken, D.T. (1968). Statistical significance in psychological research. Psychological Bulletin, 70, 151-159.

Macdonald, R.R. (1984). Algorithm AS 201: Combined significance test of differences between conditions and ordinal predictions. Applied Statistics, 33, 245-248.

Macdonald, R.R., & Smith, P.T. (1983). Testing for differences between means with ordered alternatives. British Journal of Mathematical and Statistical Psychology, 36, 22-35.

Mahoney, M.J. (1978). Experimental methods and outcome evaluation. Journal of Consulting and Clinical Psychology, 64, 660-672.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209-220.

Mantel, N., & Valand, R.S. (1970). A technique of nonparametric multivariate analysis. Biometrics, 27, 547-558.

Mantel, N. (1983). Ordered alternatives and the 1 1/2-tail test. The American Statistician, 37, 225-228.

Marascuilo, L.A., & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, Cal.: Brooks/Cole.

Marcus, R. (1976). The powers of some tests of the equality of normal means against an ordered alternative. Biometrika, 63, 177-183.

Marcus, R., & Talpaz, H. (1983). On testing homogeneity of t normal means against ordered alternatives in r groups. Communications in Statistics - Theory and Methods, A12, 2897-2902.

Marcus, R., Peritz, E., & Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. Biometrika, 63, 655-660.

Marcuson, R., & Nordbrook, E. (1981). A multivariate permutation test for the analysis of arbitrarily censored survival data. Biometrical Journal, 23, 461-465.

Mardia, K.V. (1967). A non-parametric test for the bivariate two-sample location problem. Journal of the Royal Statistical Society, Series B, 29, 320-342.

Mardia, K.V. (1971). The effect of nonnormlity on some multivariate tests and robustness to nonnormality in the linear model. Biometrika, 58, 105-121.

Maritz, J.S. (1981). Distribution-free statistical methods. London: Chapman & Hall.

Marriott, F.H.C. (1979). Barnard's Monte Carlo tests: How many simulations. Applied Statistics, 28, 75-77.

Marsaglia, G. (1972). The structure of linear congruential sequences. In S.K. Zaremba (Ed.). Application of Number Theory to Numerical Analysis. New York: Academic Press.

Marsaglia, G. (1985). A current view of random number generators. In L. Billard (Ed.), Computer signs and statistics: The Interface. (pp. 3-10). Amsterdam: Elsevier Science Publishers.

Matts, J.P., & McHugh, R.B. (1978). Analysis of accrual randomized clinical trials with balanced groups in strata. Journal of Chronical Diseases, 31, 725-740.

Matts, J.P., & McHugh, R.B. (1983). Conditional Markov chain design for accrual clinical trials. Biometrical Journal, 25, 563-577.

Maurer, W. (1981). Multiple nichtparametrische Vergleiche: Multivariate Verfahren. In U. Ferner (Hrsg.), Simultane Hypothesenprüfungen. Bad Ischl; Biometrisches Seminar der Region österreich-Schweiz der Internationalen Biometrischen Gesellschaft.

Maurer, W. (ohne Jahr). Multiple nichtparametrische Vergleiche: Multivariate Verfahren. Unveröffentlichtes Manuskript, Basel.

McCarthy, M.D. (1939). On the application of the z-test to randomized blocks. The Annals of Mathematical Statistics, 10, 337-359.

McHugh, R.B., & Matts, J.P. (1983). Post-stratification in the randomized clinical trial. Biometrics, 39, 217-225.

Meehl, P.E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science, 34, 103-115.

Meehl, P.E. (1978). Theoretical risks and tabular asteriks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Mehra, K.L., & Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments. The Annals of Mathematical Statistics, 38, 90-107.

Mehra, K.L., & Sen, P.K. (1969). On a class of conditionally distribution-free tests for interactions in factorial experiments. The Annals of Mathematical Statistics, 40, 658-664.

Meredith, W.M., Frederiksen, C.H., & McLaughlin, D.H. (1974). Statistics and data analysis. Annual Review of Psychology, 25, 453-505.

Meschkowski, H. (1968). Wahrscheinlichkeitsrechnung. Mannheim: Bibliographisches Institut.

Mielke, P.W. (1978). Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique. Biometrics, 34, 277-282.

Mielke, P.W. (1979a). Some parametric, nonparametric and permutation inference procedures resulting from weather modification experiments. Communications in Statistics - Theory and Methods, A8, 1083-1096.

Mielke, P.W. (1979b). On asymptotic non-normality of null distributions of MRPP statistics. Communications in Statistics - Theory and Methods, A8, 1541-1550.

Mielke, P.W., Berry, K.J., & Johnson, E.S. (1976). Multi-response procedures for a priori classifications. Communications in Statistics - Theory and Methods, A5, 1409-1424.

Mielke, P.W., & Berry, K.J. (1983). Asymptotic clarifications, generalizations and concerns regarding an extended class of matched pairs tests based on powers of ranks. Psychometrika, 48, 483-485.

Mielke, P.W., Berry, K.J., & Medina, J.G. (1982). Climax I and II: Distortion resistant residual analyses. Journal of Applied Meteorology, 21, 788-793.

Mielke, P.W., Berry, K.J., Brockwell, P.J., & Williams, J.S. (1981). A class of nonparametric tests based on multiresponse permutation procedures. Biometrika, 68, 720-724.

Mielke, P.W., & Iyer, H.K. (1982). Permutation techniques for analyzing multi-response data from randomized block experiments. Communications in Statistics - Theory and Methods, A11, 1427-1437.

Mielke, P.W., & Sen, P.K. (1981). On asymptotic non-normal null distributions for locally most powerful rank test statistics. Communications in Statistics - Theory and Methods, A10, 1079-1094.

Mikhail, N.N. (1965). A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis. Biometrika, 52, 149-156.

Mikhail, N.N. (1985). Penrose's criterion and Pillai's statistic under non-normal theory. Communications in Statistics - Theory and Methods, A14, 3101-3106.

Miller, R.G. (1981). Simultaneous statistical inference. New York: Springer.

Mitra, S.K. (1960). On the F-test in the intrablock analysis of a balanced incomplete block design. Sankhya: The Indian Journal of Statistics, 22, 279-284.

Mitra, S.K. (1980). Generalized inverse of matrices and applications to linear models. In P.R. Krishnaiah (Ed.), Handbook of Statistics 1: Analysis of Variance (pp, 471-512). Amsterdam: North Holland.

Möbus, C., & Nagl, W. (1983). Messung, Analyse und Prognose von Veränderungen. In J. Bredenkamp & H. Feger (Hrsg), Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I: Forschungsmethoden der Psychologie. Bd. 5: Hypothesenprüfung (S. 239-470). Göttingen: Hogrefe.

Molenaar, I.W. (1985). Statistics in the social and behavioral sciences. Statistica Neerlandica, 39, 169-179.

Monlenzun, C.J., Blouin, D.C., & Malone, L.C. (1984). Contrasting split plot and repeated measures experiments and analyses. The American Statistician, 38, 21-31.

Montgomery, D.C. (1976). Design and analysis of experiments. New York: Wiley.

Moosbrugger, H., & Steyer, R. (1983). Uni-und multivariate Varianzanalyse mit festen Parametern. In J. Bredenkamp & H. Feger (Hrsg.), Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie I. Forschungsmethoden der Psychologie. Bd. 4: Strukturierung und Reduzierung von Daten (S. 154-205). Göttingen: Hogrefe.

Morrison, D.E., & Henkel, R.E. (Eds.) (1970). The significance test controversy - a reader. London: Butterworths.

Morrison, D.F. (1976). Multivariate statistical methods (2nd ed.). New York: McGraw-Hill.

Moses, L.E. (1952). Nonparametric statistics for psychological research. Psychological Bulletin, 49, 122-143.

Murphy, B.P. (1967). Some two-sample tests when the variances are unequal: a simulation study. Biometrika, 54, 679-683.

Nair, K.R. (1940). The median in tests by randomization. Sankhya: The Indian Journal of Statistics, 4, 543-550.

Nanda, D.N. (1950). Distribution of the sum of roots of a determinantal equation under a certain condition. The Annals of Mathematical Statistics, 21, 432-439.

Nelder, J.A. (1954). The interpretation of negative components of variance. Biometrika, 41, 544-548.

Nelder, J.A. (1964a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. Proceedings of the Royal Statistical Society, Series A, 283, 147-162.

Nelder, J.A. (1964b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. Proceedings of the Royal Statistical Society, Series A, 283, 163-178.

Neumann, N. (1985). Ein Shift-Algorithmus für Blockpläne. Unveröffentlichtes Manuskript. Universität Göttingen. Abteilung für Medizinische Statistik und Dokumentation, Göttingen.

Neyman, J. (1923). Sur les applications de la theorie des probabilites aux experience agricoles: Essay de principes. Roczniki Nank Polniczek, 10, 1-51.

Neyman, J. (1967). R.A. Fisher (1890-1962): An appreciation. Science, 156, 1456-1462.

Neyman, J. (1976). Tests of statistical hypotheses and their use in studies of natural phenomena. Communications in Statistics - Theory and Methods, A5, 737-751.

Neyman, J., with cooperation of K. Iwaskiewicz, & S. Kolodziejczyk (1935). Statistical problems in agricultural experimentation. Journal of the Royal Statistical Society, 2 (Suppl.), 107-135.

Neyman, J., & Pearson, K. (1933a). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London, A231, 289-337.

Neyman, J., & Pearson, E.S. (1933b). The testing of statistical hypotheses in relation to probabilities a priori. Proceedings of the Cambridge Philosophical Society, 29, 492-510.

Nijenhuis, A., & Wilf, H.S. (1975). Combinatorial algorithms. New York: Academic Press.

Nowak, H. (1977). Tests in linearen Modellen mit Resteffekten. Dissertation, Mathematisch-naturwissenschaftliche Fakultät, RWTH Aachen, Aachen.

O'Brien, P.N., Parente, F.J., & Schmitt, C.J. (1982). A Monte Carlo study on the robustness of four MANOVA criterion tests. Journal of Statistical Computation and Simulation, 15, 183-192.

Oetting, E.R. (1965). A comment on approximate randomization tests. The Journal of Psychology, 59, 307-308.

Ogawa, J. (1961). The effect of randomization on the analysis of randomized block design. Annals of the Institute of Statistical Mathematics, 13, 105-117.

Ogawa, J. (1963). On the null-distribution of the F-statistic in a randomized balanced incomplete block design under the Neyman model. The Annals of Mathematical Statistics, 34, 1558-1568.

Ogawa, J. (1974). Statistical theory of the analysis of experimental designs. New York: Marcel Dekker.

Ogawa, J. (1980). Experimental randomization in statistical inference. In K. Matusita (Ed.), Recent developments in statistical inference and data analysis (pp. 237-246). Amsterdam: North Holland.

Ogawa, J. (1982). The Fisher experimental randomization. Journal of the Japanese Statistical Society, 12, 1-24.

Okuno, T. (1958). On the design and analysis of field experiments. II. - Split-plot and split-block designs. Bulletin of the National Institute of Agricultural Sciences, Series A6, 140-146.

Okuno, T. (1962). Mathematical models and robust criteria in the study of analysis of variance. Bulletin of the National Institute of Agricultural Sciences, Series A9, 154-204.

Oldenbürger, H.-A. (1981). Methodenheuristische Überlegungen und Untersuchungen zur "Erhebung" und Repräsentation kognitiver Strukturen. Dissertation; Mathematisch-Naturwissenschaftliche Fakultät, Universität Göttingen, Göttingen.

Olson, C.L. (1974). Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statisticsl Association, 69, 894-908.

Olson, C.L. (1975). A Monte Carlo investigation of the robustness of multivariate analysis of variance (Doctoral dissertation, University of Toronto, 1973). Dissertation Abstracts International, 35, 6106B (Microfilm, National Library of Canada, Ottawa).

Olson, C.L. (1976). On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 83, 579-586.

O'Neill, R., & Wetherill, G.B. (1971). The present state of multiple comparison methods. (with discussion). Journal of the Royal Statistical Society, Series B, 33, 218-250.

O'Reilley, F., & Mielke, P.W. (1980). Asymptotic normality of MRPP statistics from invariance principles of U-statistics. Communications in Statistics - Theory and Methods, A9, 629-637.

Pagano, M., & Tritchler, D. (1983). On obtaining permutation distributions in polynomial time. Journal of the American Statistical Association, 83, 435-440.

Pearson, E.S. (1937). Some aspects of the problem of randomization. Biometrika, 29, 53-64.

Pedhazur, E.J. (1977). Coding subjects in repeated measures designs. Psychological Bulletin, 84, 298-305.

Peritz, E. (1982). Exact tests for matched pairs: Studies with covariates. Communications in Statistics - Theory and Methods, A11, 2157-2167.

Petermann, F., & Hehl, F.J. (1979). Einzelfallanalyse. München: Kohlhammer.

Peto, R. (1973). Algorithm AS 56: Permutational significance testing. Applied Statistics, 22, 112-118.

Petrinovich, L.F., & Hardyck, C.D. (1969). Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 71, 43-54.

Petrondas, D.A., & Gabriel, K.R. (1981). Multiple comparisons by re-randomization tests. Technical Report 81/7, Department of Statistics and Division of Biostatistics, University of Rochester, New York, January, 1982.

Petrondas, D.A., & Gabriel, K.R. (1983). Multiple comparisons by rerandomization tests. Journal of the American Statistical Association, 78, 949-957.

Picard, R. (1978). Randomization and design: II. In S.E. Fienberg, & D.V. Hinckley (Eds.), R.A. Fisher: An appreciation (pp. 46-58). Berlin: Springer.

Pillai, K.C.S. (1955). Some new test criteria in multivariate analysis. The Annals of Mathematical Statistics, 26, 117-121.

Pillai, K.C.S. (1976). Distributions of characteristic roots in multivariate analysis. Part I. Null distributions. The Canadian Journal of Statistics, Sections A & B: Theory and Methods, 4, 157-184.

Pillai, K.C.S. (1977). Distributions of characteristic roots in multivariate analysis. Part II: Non-null distributions. The Canadian Journal of Statistics, Sections A & B: Theory and Methods, 5, 1-62.

Pincus, R., & Majumdar, D. (1981). Monte Carlo studies for the power of a one side test for homogeneity of means in an unbalanced one way model. Sankhya: The Indian Journal of Statistics, B43, 104-111.

Pitman, E.J.G. (1937a). Significance tests which may be applied to samples from any populations. Journal of the Royal Statistical Society, 4, (Suppl.), 119-130.

Pitman, E.J.G. (1937b). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. Journal of the Royal Statistical Society, 4, (Suppl.), 225-232.

Pitman, E.J.G. (1938). Significance tests which may be applied to samples from any populations. III. The analysis of variance test. Biometrika, 29, 322-335.

Popper, K.R. (1974). Objektive Erkenntnis: Ein evolutionärer Entwurf (2. Aufl.), Hamburg: Hoffmann und Campe.

Popper, K.R. (1976). Logik der Forschung (6. Aufl.), Tübingen: J.C.B. Mohr.

Pratt, J.W. (1976). A discussion of the question: For what use are tests of hypotheses and tests of significance. Communications in Statistics - Theory and Methods, A5, 779-787.

Pratt, J.W., & Gibbons, J.D. (1981). Concepts of nonparametric theory. New York: Springer.

Preece, D.A. (1983). Discussion of the paper by Wilkinson et al. Journal of the Royal Statistical Society, Series B, 45, 202-211.

Puri, M.L., & Sen, P.K. (1967). On some optimum nonparametric procedures in two-way layouts. Journal of the American Statistical Association, 62, 1214-1229.

Puri, M.L., & Sen, P.K. (1971). Nonparametric methods in multivariate analysis. New York: Wiley.

Pyhel, N. (ohne Jahr). Kombinatorische Algorithmen für Anwendungen in der Statistik. Unveröffentlichtes Manuskript, Abteilung Medizinische Statistik und Dokumentation, RWTH Aachen, Aachen.

Pyhel, N. (1978a). Beiträge zur Theorie der Permutationstests mit Anwendung in einem linearen Modell. Dissertation, Mathematisch-Naturwissenschaftliche Fakultät, RWTH Aachen, Aachen.

Pyhel, N. (1978b). On permutationally invariant distributions and exactly distribution-free tests of linear hypotheses. Unveröffentlichtes Manuskript. Abteilung für Medizinische Statistik und Dokumentation, RWTH Aachen, Aachen.

Pyhel, N. (1980a). Distribution-free r sample tests for the hypothesis of parallelism of response profiles. Biometrical Journal, 22, 703-714.

Pyhel, N. (1980b). Zur Anwendung von Permutationstests in mehrfaktoriellen Versuchsplänen. In W. Köpcke, & K. Überla (Hrsg.), Biometrie - heute und morgen. Interregionales Biometrisches Kolloquium (S. 350-358). Berlin: Springer.

Ramsey, P.H. (1978). Power differences between pairwise multiple comparisons. (with comment by K.R. Gabriel and rejoinder). Journal of the American Statistical Association, 73, 479-487.

Randles, R.H., & Wolfe, D.A. (1979). Introduction to the theory of nonparametric statistics. New York: Wiley.

Rao, C.R. (1959). Expected values of mean squares in the analysis of incomplete block experiments and some comments based on them. Sankhya: The Indian Journal of Statistics, 21, 327-336.

Rao, C.R. (1965). Linear statistical inference and its applications. New York: Wiley.

Rao, C.R., & Mitra, S.K. (1971). Generalized inverse of matrices and its applications. New York: Wiley.

Rao, P.V. (1963). The F-test in the intrablock analysis of a class of PBIB designs. Annals of the Institute of Statistics and Mathematics, 15, 25-36.

Recchia, M., & Rocchetti, M. (1982). The simulated randomization test. Computer Programs in Biomedicine, 15, 111-116.

Ripley, B.D. (1983). Computer generation of random variables: A tutorial. International Statistical Review, 51, 301-319.

Robertson, T., & Wright, F.T. (1982). On measuring the conformity of a parameter set to a trend, with applications. The Annals of Statistics, 10, 1234-1245.

Roberts, H.V. (1976). For what use are tests of hypotheses and tests of significance. Communications in Statistics - Theory and Methods, A5, 753-761.

Robinson, J. (1972). The randomization model for incomplete block designs. The Annals of Mathematical Statistics, 43, 480-489.

Robinson, J. (1973a). The large-sample power of permutation tests for randomization models. The Annals of Statistics, 1, 291-296.

Robinson, J. (1973b). The analysis of covariance under a randomization model. Journal of the Royal Statistical Society, Series B, 35, 368-376.

Robinson, J. (1978). An asymptotic expansion for samples from a finite population. The Annals of Statistics, 6, 1005-1011.

Robinson, J. (1979). Optimal tests of significance. Australian Journal of Statistics, 21, 301-310.

Robinson, J. (1980). An asymptotic expansion for permutation tests with several samples. The Annals of Statistics, 8, 851-864.

Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. Journal of the Royal Statistical Society, Series B, 44, 91-101.

Robinson, J. (1983). Approximations to some test statistics for permutation tests in a completely randomized design. Australian Journal of Statistics, 25, 358-369.

Roebruck, P. (1978). Über optimale Tests in linearen Modellen mit gemischten Effekten. Dissertation, Mathematisch-Naturwissenschaftliche Fakultät, RWTH Aachen, Aachen.

Roebruck, P. (1979). Lineare Modelle mit gemischten Effekten. Unveröffentlichtes Vorlesungsskript. Abteilung für Medizinische Statustik und Dokumentation, RWTH Aachen, Aachen.

Roebruck, P. (1982a). Canonical forms and tests of hypotheses - Part I: The general univariate mixed model. Statistica Neerlandica, 36, 63-74.

Roebruck, P. (1982b). Canonical forms and tests of hypotheses - Part II: Multivariate mixed linear models. Statistica Neerlandica, 36, 75-80.

Roebruck, P. (1983). Zur Optimalität des F-Tests für lineare Hypothesen über lineare Modelle mit gemischten Effekten. Metrika, 30, 85-91.

Roebruck, P., & Willmes, K. (1983). Nichtnegative Varianzkomponentenschätzungen in ANOVA-Modellen mit gemischten Effekten und ihre Anwendung auf die Generalisierbarkeitstheorie. Diagnostica, 29, 1-25.

Rogan, J.C., Keselman, H.J., & Mendoza, J.L. (1979). Analysis of repeated measurements. British Journal of Mathematical and Statistical Psychology, 32, 269-286.

Rosenbaum, P.R. (1984). Conditional permutation tests and the propensity score in observational studies. Journal of the American Statistical Association, 79, 565-574.

Roth, A.J. (1983). Robust trend tests derived and simulated: Analogs of the Welch and Brown-Forsythe tests. Journal of the American Statistical Association, 78, 972-980.

Rothe, G. (1986). Linear trend test versus global test: A comparison. Statistica Neerlandica, 40, 1-11.

Rouanet, H., Bernard, J.-M., & Lecoutre, B. (1986). Nonprobabilistic statistical inference: A set-theoretic approach. The American Statistician, 40, 60-65.

Roux, C.Z. (1982). The analysis of covariance in randomization theory. South African Statistical Journal, 16, 1-23.

Royen, T. (1978). Randomisierungstests zum Vergleich verbundener Stichproben. EDV in Medizin und Biologie, 9, 104-106.

Rubin, D.B. (1978). Bayesian inference for casual effects: The role of randomization. The Annals of Statistics, 6, 34-58.

Rubinstein, R.Y. (1981). Simulation and the Monte Carlo method. New York: Wiley.

Ryan, T.A. (1959). Multiple comparisons in psychological research. Psychological Bulletin, 56, 26-47.

Ryan, T.A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. Psychological Bulletin, 57, 318-328.

Sachdeva, D. (1973). Estimating strength of relationship in multivariate analysis of variance. Educational and Psychological Measurement, 33, 627-631.

Sahner, H. (1979). Veröffentlichte empirische Sozialforschung: Eine Kumulation von Artefakten? Zeitschrift für Soziologie, 8, 267-278.

Savage, L.J. (1976). On rereading R.A. Fisher. The Annals of Statistics, 4, 441-500.

Scariano, S.M., & Davenport, J.M. (1984). Corrected F-tests in the general linear model. Communications in Statistics - Theory and Methods, A13, 3155-3172.

Searle, S.R. (1971). Linear models. New York: Wiley.

Searle, S.R. (1982). Matrix algebra useful for statistics. New York: Wiley.

Sen, P.K. (1965). On some permutation tests based on U-statistics. Calcutta Statistical Association Bulletin, 14, 106-126.

Sen, P.K. (1967). On some multisample permutation tests based on a class of U-statistics. Journal of the American Statistical Association, 62, 1201-1213.

Sen, P.K. (1968). On a class of aligned rank order tests in two-way layouts. The Annals of Mathematical Statistics, 39, 1115-1124.

Shaffer, J.P. (1980). Control of directional errors with stagewise multiple test procedures. The Annals of Statistics, 8, 1342-1347.

Shaffer, J.P. (1984). Issues arising in multiple comparisons among populations. In M. Iosifescu (Ed.), Proceedings of the 7th Conference on Probability Theory. Bukarest: Editura Academici Republici Socialiste Romania.

Shaffer, J.P. (1986). Modified sequentially rejective multiple test procedures. Unpublished Manuscript, University of California, Department of Statistics.

Shen, C.D., & Quade, D. (1983). A randomization test for a three-period three-treatment crossover experiment. Communications in Statistics - Simulation and Computation, B12, 183-199.

Shiraishi, T.-A. (1984). Semi-aligned rank tests. Annals of the Institute of Statistics and Mathematics, A36, 463-473.

Shorack, G.R. (1967). Testing against ordered alternatives in model I analysis of variance; normal theory and nonparametric. The Annals of Mathematical Statistics, 38, 1740-1752.

Shuster, J.J. (1978). Practical permutation tests in complex experimental design. In American Statistical Association, Proceedings of the Statistical Computing Section (pp. 174-175), Washington.

Shuster, J.J., & Boyett, J.M. (1979). Nonparametric multiple comparison procedures. Journal of the American Statistical Association, 74, 379-382.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw Hill.

Siemiatycki, J. (1978). Mantel's space-time clustering statistic: Computing higher moments and a comparison of various data transforms. Journal of Statistical Computation and Simulation, 7, 13-31.

Silvey, S.D. (1953). The equivalence of asymptotic distributions under randomization and normal theories. Proceedings of the Glasgow Mathematical Association, 1, 139-147.

Silvey, S.D. (1954). The asymptotic distributions of statistics arising in certain non-parametric tests. Proceedings of the Glasgow Mathematical Association, 2, 47-51.

Silvey, S.D. (1970). Statistical inference. Harmondsworth, Middlesex: Penguin Books.

Singer, B. (1981). Review of randomization tests by E.S. Edgington. British Journal of Mathematical and Statistical Psychology, 34, 261-262.

Smith, I.L. (1972). The eta coefficient in MANOVA. Multivariate Behavioral Research, 7, 361-372.

Smith, P.T., & Macdonald, R.R. (1983). methods for incorporating ordinal information into analysis of variance: Generalizations of one-tail tests. British Journal of Mathematical and Statistical Psychology, 36, 1-21.

Smythe, R.T., & Wei, L.J. (1983). Significance tests with restricted randomization design. Biometrika, 70, 496-400.

Soms, A.P. (1977). An algorithm for the discrete Fisher's permutation test. Journal of the American Statistical Association, 72, 662-664.

Soms, A.P. (1985). Permutation tests for k-sample binomial data with comparisons of exact and approximate p-levels. Communications in Statistics - Theory and Methods, A14, 217-233.

Sonnemann, E. (1981a). Tests zum multiplen Niveau alpha. In U. Ferner, (Hrsg.), Simultane Hypothesenprüfungen. Bad Ischl, Biometrisches Seminar der Region österreich-Schweiz der Internationalen Biometrischen Gesellschaft.

Sonnemann, E. (1981b). Tests zum multiplen Niveau alpha. Vortrag auf dem Seminar der ROeS, Bad Ischl, 2. verbesserte Ausarbeitung. Unveröffentlichtes Manuskript, Universität Dortmund, Abteilung Statistik, Dortmund.

Sonnemann, E. (1982). Allgemeine Lösungen multipler Testprobleme. EDV in Medizin und Biologie, 13, 120-128.

Schatzoff, M. (1966). Sensitivity comparisons among tests of the general linear hypothesis. Journal of the American Statistical Association, 61, 415-435.

Scheffe, H. (1943a). On a measure problem arising in the theory of non-parametric tests. The Annals of Mathematical Statistics, 14, 227-233.

Scheffe, H. (1943b). Statistical inference in the non-parametric case. The Annals of Mathematical Statistics, 14, 305-332.

Scheffe, H. (1959). The analysis of variance. New York: Wiley.

Schiller, K. (1981). Multivariate Varianzanalyse: Simultane Vergleiche. In U. Ferner (Hrsg.), Simultane Hypothesenprüfungen. Bad Ischl, Biometrisches Seminar der Region österreich-Schweiz der Internationalen Biometrischen Gesellschaft.

Schrader, R.M., & McKean, J.W. (1977). Robust analysis of variance. Communications in Statistics - Theory and Methods, A6, 879-894.

Steinhorst, R.K. (1982). Resolving current controversies in analysis of variance. The American Statistician, 36, 138-139.

Stein-Zaar, G. (1983). Permutationstests in mehrfaktoriellen linearen Modellen. Diplomarbeit: Abteilung für Medizinische Statistik und Dokumentation, RWTH Aachen, Aachen.

Sterling, T.D. (1959). Publication decisions and their possible effects on inferences from tests of significance - or vice versa. Journal of the American Statistical Association, 54, 30-34.

Stevens, J.P. (1972). Global measures of association in multivariate analysis of variance. Multivariate Behavioral Research, 7, 373-378.

Stevens, J.P. (1979). Comment on Olson: Choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 86, 355-360.

Still, A.W., & White, A.P. (1981). The approximate randomization test as an alternative to the F-test in analysis of variance. British Journal of Mathematical and Statistical Psychology, 34, 243-252.

Stoline, M.R., & Mitchell, B.T. (1981). Generation of the Hunter second order Bonferroni approximation to the multivariate t and its applications (Report No. 1). Western Michigan University Center for Statistical Services.

Streitberg, B. (1986). The permutation test problem is NP-hard. Statistical Software Newsletter, 12, 18-19.

Streitberg, B., & Röhmel, J. (1983). Exakte Verteilungen für Rang- und Randomisierungstests im allgemeinen c-Stichprobenproblem. Vortrag auf der Tagung der International Biometric Society, Sektion Deutschland, März, Bad Nauheim.

Streitberg, B., & Röhmel, J. (1984). Exakte Verteilungen für Rang- und Randomisierungstests im allgemeinen c-Stichprobenproblem. Unveröffentlichtes Manuskript.

Streitberg, B., & Röhmel, J. (1986). Exact distributions for permutation and rank tests: An introduction to some recently published algorithms. Statistical Software Newsletter, 12, 10-17.

Stucky, W., & Vollmar, J. (1976). Exact probabilities for tied linear rank tests. Journal of Statistical Computation and Simulation, 5, 73-81.

Taylor, M.S., & Thompson, J.R. (1986). A data based algorithm for the generation of random vectors. Computational Statistics & Data Analysis, 4, 93-101.

Thornett, M.L. (1982a). The role of randomization in model-based inference. Australian Journal of Statistics, 24, 137-145.

Thornett, M.L. (1982b). Two views of randomization? Response to Peter Finch's comments. Australian Journal of Statistics, 24, 148-150.

Timm, N.H. (1980). Multivariate analysis of variance of repeated measurements. In P.R. Krishnaiah (Ed.), Handbook of statistics 1: Analysis of variance (pp. 41-87). Amsterdam: North Holland.

Toothaker, L.E. (1971). The effect of the joint violation of two assumptions on the size of the F-test under permutation for the randomized block design. Journal of Statistical Computation and Simulation, 1, 55-64.

Toothaker, L.E. (1972). An empirical investigation of the permutation t-test. British Journal of Mathematical and Statistical Psychology, 25, 83-94.

Tracy, D.S., & Tajuddin, J.H. (1985). Extended moment results for improving inferences based on MRPP. Communications in Statistics - Theory and Methods, A14, 1485-1496.

Tracy, D.S., & Tajuddin, J.H. (1986). Empirical power comparisons of two MRPP rank tests. Communications in Statistics - Theory and Methods, A15, 551-570.

Tritchler, D. (1984). On inverting permutation tests. Journal of the American Statistical Association, 79, 200-207.

Tsutakawa, R.K., & Yang, S.L. (1974). Permutation tests applied to antibiotic drug resistance. Journal of the American Statistical Association, 69, 87-92.

Tukey, J.W. (1960). Conclusions vs. decisions. Technometrics, 2, 423-433.

Tullock, G. (1959). Publication decisions and tests of significance - a comment. Journal of the American Statistical Association, 54, 593.

Vadiveloo, J. (1983). On the theory of modified randomization tests for nonparametric hypotheses. Communications in Statistics - Theory and Methods, A12, 1581-1596.

van Dantzig, D., & Hemelrijk, J. (1954). Statistical methods based on few assumptions. Bulletin of the International Statistical Institute, 34, 3-31.

van Elteren, P., & Verbeek, A. (1979). De correlatie - coefficient en permutatie-toetsen. Vereniging voor Statistiek Bulletin, 12, 35-44.

Victor, N. (1984). Computational statistics: Tool or Science? - Werkzeug oder Wissenschaft? Statistical Software Newsletter, 120, 105-125.

Wainer, H. (1976). Robust statistics: A survey and some prescriptions. Journal of Educational Statistics, 1, 285-312.

Wald, A., & Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. The Annals of Mathematical Statistics, 15, 358-372.

Wall, M., & Meystre, A. (1974). Testings simple hypotheses by a Monte Carlo method with sequential decision procedure. In G. Bruckmann, F. Ferschl, & L. Schmetterer (Eds.), COMPSTAT 1974: Proceedings in computational statistics (pp. 36-46). Wien: Physica.

Wehrhahn, K. (1986). Properties of some distributions in randomization theory. Journal of Statistical Planning and Inference, 13, 231-237.

Wehrhahn, K.H., & Ogawa, J. (1982). Randomization in a complete block design with few blocks. Journal of Statistical Planning and Inference, 6, 33-45.

Wei, L.J. (1983). Tests for interchangeability with incomplete paired observations. Journal of the American Statistical Association, 78, 725-729.

Wei, L.J., Smythe, R.T., & Smith, R.L. (1986). K-treatment comparisons with restricted randomization rules in clinical trials. The Annals of Statistics, 14, 265-274.

Welch, B.L. (1937). On the z-test in randomized blocks and Latin squares. Biometrika, 29, 21-52.

Welsch, R.E. (1977). Stepwise multiple comparison procedures. Journal of the American Statistical Association, 72, 566-575.

Westermann, R. (1982). Zur Wahl der Fehlerwahrscheinlichkeiten bei mehrfachen Signifikanztests. In Lüer, G. (Hrg.), Bericht über den 33. Kongress der Deutschen Gesellschaft für Psychologie, Bd. 1 (S. 193-197), Göttingen: Hogrefe.

Westermann, R., & Hager, W. (1983a). On severe tests of trend hypotheses in psychology. The Psychological Record, 33, 201-211.

Westermann, R., & Hager, W. (1983b). The relative importance of low significance level and high power in multiple tests of significance. Perceptual and Motor Skills, 56, 407-413.

White, R.F. (1975). Randomization and the analysis of variance. Biometrics, 31, 552-572.

Wichmann, B.A., & Hill, I.D. (1982). An efficient and portable pseudo-random number generator. Applied Statistics, 31, 188-190.

Wilk, M.B. (1955). The randomization analysis of a generalized randomized block design. Biometrika, 42, 70-79.

Wilk, M.B. (1957). Non-additivities in a Latin square design. Journal of the American Statistical Association, 52, 218-236.

Wilk, M.B., & Kempthorne, O. (1955). Fixed, mixed, and random models. Journal of the American Statistical Association, 50, 1144-1167.

Wilk, M.B., & Kempthorne, O. (1956). Some aspects of the analysis of factorial experiments in a completely randomized design. The Annals of Mathematical Statistics, 27, 950-985.

Wilk, M.B., & Kempthorne, O. (1957). Non-additivities in a Latin square design. Journal of the American Statistical Association, 52, 218-236.

Wilkinson, L. (1975). Response variate hypotheses in the multivariate analysis of variance. Psychological Bulletin, 82, 408-412.

Wilks, S.S. (1962). Mathematical statistics. New York: Wiley.

Williams, D.A. (1971). A test for differences between treatment means when several dose levels are compared with a zero control. Biometrics, 27, 103-117.

Willmes, K. (1981a). Permutationstests als Alternative zur MANOVA. Vortrag: 27. Biometrisches Kolloquium der Deutschen Region der Internationalen Biometrischen Gesellschaft, Bad Nauheim.

Willmes, K. (1981b). Permutationstests für lineare Hypothesen. Vortrag, 23. Tagung experimentell arbeitender Psychologen, Berlin.

Willmes, K. (1982). A comparison between the Lehmacher & Wall rank tests and Pyhel's permutation test for the analysis of r independent samples of response curves. Biometrical Journal, 24, 717-722.

Willmes, K., & Pyhel, N. (1981). Permutationstests als Alternative zur Varianzanalyse - Der Split-Plot Versuchsplan. Zeitschrift für Sozialpsychologie, 12, 186-198.

Winer, B.J. (1971) Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wirth, N. (1983). Algorithmen und Datenstrukturen. Stuttgart: Teubner.

Witte, E.H. (1980). Signifikanztest und statistische Inferenz. Stuttgart: Enke.

Witting, H. (1969). Mathematische Statistik. Stuttgart: Teubner.

Witting, H. (1985). Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang. Stuttgart: Teubner.

Witting, H., & Nölle, G. (1970). Angewandte mathematische Statistik. Stuttgart: Teubner.

Woodward, J.A., & Overall, J.E. (1975). Multivariate analysis of variance by multiple regression methods.

Worsley, K.J. (1982). An improved Bonferroni inequality and applications. Biometrika, 69, 297-302.

Wu, C.F. (1981).On the robustness and efficiency of some randomized designs. The Annals of Statistics, 9, 1168-1177.

Yates, F. (1964). Sir Ronald Fisher and the design of experiments. Biometrics, 20, 307-321.

Yates, F. (1975). The early history of experimental design. In J.N. Srivastava (Ed.),A survey of statistical design and linear models (pp. 581-512). North Holland Publishing Company.

Zerbe, G.O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. Journal of the American Statistical Association, 74, 215-257.

Zerbe, G.O., & Walker, S.H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design matrices. Biometrics, 33, 653-657.

Zimmermann, H. (1985a). Exact calculation of permutational distributions for two dependent samples I. Biometrical Journal, 27, 349-352.

Zimmermann, H. (1985b). Exact calculation of permutational distribution for two independent samples. Biometrical Journal, 27, 431-434.

Zyskind, G. (1962). On structure, relation, Sigma, and expectation of mean squares. Sankhya: The Indian Journal of Statistics, Series A, 24, 115-148.

Zyskind, G. (1963). Some consequences of randomization in a generalization of the balanced incomplete block design. The Annals of Mathematical Statistics, 34, 1569-1581.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. The Annals of Mathematical Statistics, 38, 1092-1109.

Zyskind, G. (1969). Parametric augmentations and error structures under which certain simple least squares and analysis of variance procedures are also best. Journal of the American Statistical Association, 64, 1353-1368.

Zyskind, G. (1975). Error structures, projections and conditional inverses in linear model theory. In J.N. Srivastava (Ed.), A survey of statistical design and linear models (pp. 647-663). New York: North Holland/American Elsevier.

Überblick

A1 Bezeichnungen

A2 Eigenschaften verschiedener Matrizenoperationen

A3 Eigenschaften von Permutationsgruppen

A4 Lineare Unterräume und Projektionen

B1 FORTRAN-Version des Shift-Algorithmus von Streitberg & Röhmel für das unverbundene Zweistichprobenproblem

B2 FORTRAN-Version des Shift-Algorithmus von Streitberg & Röhmel für das Einstichproben- bzw. verbundene Zweistichprobenproblem

B3 FORTRAN-Listing der sequentiellen multiplen Testprozedur für die abhängigen Variablen in einem multivariaten Testproblem: Subroutine SEQMTP

B4 FORTRAN-Listing des Programmes für den Test auf Wechselwirkungen im multivariaten Split-Plot Versuchsplan

B5 Verzeichnis der vorhandenen FORTRAN-Programme

ANHANG A

A1: BEZEICHNUNGEN

Matrizen und Vektoren

$A_{n\times m}$	(n×m)-dimensionale Matrix (oft ohne Angabe der Dimensionen)
I_n	(n×n)-dimensionale Einheitsmatrix
J_n	(n×n)-dimensionale Matrix mit Einsen
E_n	$I_n - 1/nJ_n$
1_n	n-dimensionaler Spaltenvektor mit Einsen
0_n	n-dimensionaler Spaltenvektor mit Nullen (bzw. Nullmatrix)
X	oft: (n×m)-dimensionale Design-Matrix in einem linearen Modell
P	oft: (n×n)-dimensionale Permutationsmatrix zu einer Permutation
K	oft: Transformationsmatrix und die durch sie erzeugte Transformation: $R^n \to R^n$
$((a_{ij}))_{n\times m}$	(n×m)-dimensionale Matrix mit Elementen a_{ij}, i=1,..,n;j=1,..,m
A'	Transponierte einer Matrix
A^{-1}	Inverse einer quadratischen Matrix
A^-	Verallgemeinerte Inverse einer (quadratischen) Matrix oft: Moore-Penrose Inverse (Mitra,1980; Searle,1982)
rg(A)	Rang der Matrix A
sp(A)	Spur der quadratischen Matrix A
det(A)	Determinante der quadratischen Matrix A
λ_i, θ_i	oft: reelle Eigenwerte einer quadratischen Matrix
A⊗B	Kronecker-Produkt zweier Matrizen A,B
A⊕B	Kronecker-Summe zweier Matrizen A,B
$\bigoplus_{i=1}^{n}$	Kronecker-Summe von n Matrizen A_i, i=1,..,n
A∔B	Spaltensumme zweier Matrizen A,B
A⊟B	Zeilensumme zweier Matrizen A,B
A*B	Hadamard-Produkt zweier Matrizen A,B
ab'	Dyadischen Produkt zweier Spaltenvektoren a,b

M_n	$(I_{n-1}\|-1_{n-1})'$
S_h, S_e	Summe von Quadraten und Kreuzprodukten für Hypothesen- und Fehlerterm in der MANOVA
$\Sigma, \mathfrak{Z}, V_n$	oft: Varianz-Kovarianz Matrix für einen n-dimensionalen Zufallsvektor
$((\sigma^*_{jj'}))_{n\times n}$	oft: (verallgemeinerte) Inverse einer Varianz-Kovarianzmatrix $((\sigma_{jj'}))_{n\times n}$
$\mu, \theta, \Delta, \alpha, \beta, \gamma$	Parametervektoren

Gruppen

G	(allgemeine) Gruppe, oft Gruppe linearer Transformationen
S_n	Symmetrische Gruppe der Ordnung n, Gruppe aller Permutationen von n Elementen (speziell von n reellen Zahlen)
Q_n, Q_n^*	(echte) Untergruppen von S_n
$\{1_n\}$	Permutationsgruppe bestehend aus der identischen Permutation
π, ρ	Elemente von S_n bzw. deren zugehörige Transformationsmatrix
$\tilde{S}_n$	Symmetrische Gruppe der n-dimensionalen Vektorpermutationen isomorph zu S_n
$\tilde{Q}_n, \tilde{Q}_n^*$	Echte Untergruppen von $\tilde{S}_n$
$S_n(K, Q_n)$	Gruppe der K-Q_n-zulässigen Permutationen mit K Transformationsmatrix: $\{\pi\in S_n \mid \pi K = K\rho, \rho\in Q_n\}$
$S_n(K)$	$\{\pi\in S_n \mid \pi K = K\pi\}$,ebenfalls Untergruppe von S_n
$S_k \otimes S_r$	Kartesisches Produkt von S_k mit S_r
$S_k \wr S_r$	Kranzprodukt von S_k mit S_r ($S_r[S_k]$ bei Pyhel,1978a)
$\overset{r}{\underset{i=1}{\times}} S_k$	r-fache Kopie von S_k
$\|G\|$	Anzahl der Elemente einer endlichen Gruppe

Räume

R^n	n-dimensionaler euklidischer Raum
U, U_i, U^H, U^A	Lineare Unterräume des R^n
$\bigoplus_{i=1}^{n} U_i$	Direkte Summe von Unterräumen
$U \setminus U^H$	Raum U ohne U^H
$\complement_\perp U$	Orthogonales Komplement von U in R^n
dim(U)	Dimension des Unterraumes U

Abbildungen

f,g	Allgemeine Abbildungen
f_H	Orthogonale Projektion $R^n \to U^H$
f_H^*	Orthogonale Projektion $R^n \to \complement_\perp U$
id	Identische Abbildung
π, ρ, π^+	Permutationen
ϕ	Testfunktion
T	oft: Transformationen einer Zufallsvariablen Y: T(Y)
S(Y), Q(Y), L_n	Teststatistiken
F(Y)	oft: Teststatistik mit F-Verteilung
$\|\|x\|\|$	Norm von $x \in R^n$
$\|x\|$	Betrag von $x \in R^1$
[x]	Größte ganze Zahl kleiner oder gleich $x \in R^1$

Mengen

$I = \{1, .., n\}$	Indexmenge
$\|I\|$	Mächtigkeit einer endlichen Menge
$A \setminus B$	Differenzmenge

Zufallsvariablen

Y,Z,ϵ	Ein- oder mehrdimensionale Zufallsvariablen
$\delta_{i\nu}^{j}$	oft: Design-Zufallsvariablen in Randomisierungsmodellen
y,z	Realisationen ein- oder mehrdimensionaler Zufallsvariablen
$E(Y),Var(Y)$	Erwartungswert, Varianz der eindimensionalen Zufallsvariablen bzw. Erwartungswertvektor, Varianz-Kovarianz Matrix des mehrdimensionalen Zufallsvektors
$Cov(X,Y)$	Kovarianz der Zufallsvariablen X,Y
$P_n, P_{n.}$	Permutationsmaße
$E_{P_n}(Y)$	Erwartungswert unter dem Permutationsmaß P_n (analog für Varianz und Kovarianz)
F,F^{ϵ}	Verteilungsfunktion einer ein- oder mehrdimensionalen Zufallsvariablen bzw. der Fehlerzufallsvariablen in einem linearen Modell
F^T	Durch die Abbildung T induzierte Verteilungsfunktion einer ein- oder mehrdimensionalen Zufallsvariablen
$NV(\mu,\sigma^2)$	Univariate Normalverteilung mit Erwartungswert μ und Varianz σ^2
$NV_p(\mu,\Sigma)$	p-variate Normalverteilung mit Erwartungswert μ und Varianz-Kovarianz Matrix Σ
χ_f^2	Zentrale chi-Quadrat Verteilung mit f Freiheitsgraden

Sonstiges

$n!$	n-Fakultät: Produkt der Zahlen $1\cdot2\cdot\ldots\cdot n$
$\sum_{i=1}^{n} a_i, \prod_{i=1}^{n} a_i$	Summe bzw. Produkt der Elemente $a_1,..,a_n$ mit Laufindex i
$\lim_{n\to\infty} a_n$	Grenzwert einer Folge von Elementen a_n, n natürliche Zahl
$\delta_{ii'}$	Kronecker-Symbol
H: , A:	Nullhypothese und Alternativhypothese
H_o:, H_1:	
$\{H_i\}$	Hypothesenfamilie mit Elementen $i\in I$
$\cap_i H_i$, $\cup_i H_i$	Durchschnitt, Vereinigung von Hypothesen
$A\subset B$	A ist echte oder unechte Teilmenge von B

$\hat{\gamma}$	oft: Kleinster-Quadrat Schätzer des Parameters γ
$\alpha_{j(i)}$	Faktor α ist unter einen Faktor, dessen Stufen mit i bezeichnet sind, genestet
$\bar{Y}_{ij.}$	Mittelung bezüglich des (der) durch . ersetzten Index (Indizes)
$x_{(1)},\ldots,x_{(n)}$	Geordnete Werte $x_1,\ldots,x_n$

Versuchspläne

CR-k,pCR-k	Einfaktorieller vollständiger Zufallsplan mit k Faktorstufen für eine uni- bzw. p-variate abhängige Variable
RB-k	Einfaktorieller vollständiger Blockplan
BIBD	Ausgewogener, unvollständiger Blockplan
CRH-r(c)	Einfach hierarchischer zweifaktorieller Zufallsplan jeweils c Stufen eines Faktors genestet unter jede der r Stufen eines anderen Faktors
CRH-r(c)(s)	Zweifach hierarchischer dreifaktorieller Zufallsplan
CRH-(rc)s	Partiell hierarchischer, dreifaktorieller Zufallsplan ein Faktor genestet unter zwei vollständig gekreuzte Faktoren
CRH-r(c)s	Partiell hierarchischer dreifaktorieller Zufallsplan zwei vollständig gekreuzte Faktoren, ein dritter Faktor genestet unter den ersten den beiden gekreuzten Faktoren
LS-k	Lateinisches Quadrat mit k Stufen
CRF-rc..	Zweifaktorieller, dreifaktorieller,.. vollständiger Zufallsplan
RBF-rc..	Zweifaktorieller, dreifaktorieller,.. vollständiger Blockplan
SPF-r.c	Split-Plot Plan, wiederholte Messungen für den zweiten Faktor mit c Stufen
SPF-rc.s	Split-Plot-Plot Plan, wiederholte Messungen für den dritten Faktor mit s Stufen
SPF-r.cs	Split-Split-Plot Plan, wiederholte Messungen für die beiden letzten, vollständig gekreuzten Faktoren

A2: EIGENSCHAFTEN VERSCHIEDENER MATRIZENOPERATIONEN

Die Dimensionen der Matrizen seien so beschaffen, daß die auftretenden Operationen definiert sind (vgl. Rao & Mitra,1971; Searle,1971,1982)

1. Kronecker-Produkt (Direktes Produkt)

$A_{n\times m} \otimes B_{p\times q} := ((a_{ij}\cdot B))_{np\times mq}$

(1) $(A\otimes B)' = A'\otimes B'$

(2) $(A\otimes B)\cdot(C\otimes D) = A\cdot C \otimes B\cdot D$

(3) $a(A\otimes B) = (aA)\otimes B = A\otimes(aB) \quad , a\in R^1$

(4) $(A+B)\otimes C = (A\otimes C)+(B\otimes C)$

$A\otimes(B+C) = (A\otimes B)+(A\otimes C)$

(5) $(A\otimes B)\otimes C = A\otimes(B\otimes C)$

(6) $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$; A,B invertierbar(regulär)

$(A\otimes B)^{-} = A^{-}\otimes B^{-}$

(7) $rg(A\otimes B) = rg(A)\cdot rg(B)$

$sp(A\otimes B) = sp(A)\cdot sp(B)$

(8) $I_n\otimes I_m = I_{nm} \qquad J_n\otimes J_m = J_{nm}$

2. Kronecker-Summe (Direkte Summe)

$$A_{n\times m} \oplus B_{p\times q} = \begin{pmatrix} A & O \\ O & B \end{pmatrix}_{(n+p)\times(m+q)}$$

(1) $(A\oplus B)' = A'\oplus B'$

(2) $(A\oplus B)\cdot(C\oplus D) = A\cdot C \oplus B\cdot D$

(3) $(A\oplus B)+(C\oplus D) = (A+C)\oplus(B+D)$

(4) $(A\oplus B)^{-1} = A^{-1}\oplus B^{-1}$; A,B invertierbar

(5) $rg(A\oplus B) = rg(A) + rg(B)$

$sp(A\oplus B) = sp(A) + sp(B)$; A,B quadratisch

(6) $I_n\oplus I_m = I_{nm}$

(7) $\bigoplus_{i=1}^{k} A = I_k\otimes A$

3. Idempotente Matrizen

Eine quadratische Matrix A heißt idempotent, falls

$$A \cdot A = A \quad .$$

Für eine idempotente Matrix A gilt: $sp(A) = rg(A)$

Spezielle idempotente Matrizen:

I_n $rg(I_n)=n$; I_n ist die einzige idempotente Matrix von vollem Rang.

$1/nJ_n$ $rg(1/nJ_n) = 1$; E_n $rg(E_n) = n-1$

4. (Verallgemeinerte) Inverse (vgl. Rao & Mitra,1971; Searle,1982)

A^{-1} heißt Inverse einer quadratischen Matrix, falls:

$$A^{-1} \cdot A = A \cdot A^{-1} = I$$

$A^-_{n\times m}$ heißt eine verallgemeinerte Inverse der Matrix $A_{n\times m}$, falls:

$$A \cdot A^- \cdot A = A$$

Eine verallgemeinerte Inverse heißt Moore-Penrose Inverse A^+ , falls:

$A \cdot A^+ \cdot A = A$, $A^+ \cdot A \cdot A^+ = A^+$; $A \cdot A^+$ und $A^+ \cdot A$ sind symmetrisch .

A^+ ist eindeutig. In dieser Arbeit wird in der Bezeichnungsweise nicht zwischen A^- und A^+ unterschieden.

Die eindeutige verallgemeinerte Inverse einer idempotenten Matrix ist diese Matrix selbst.

5. Spaltensumme, Zeilensumme

$$A_{n\times m} \boxplus B_{p\times m} := \begin{pmatrix} A \\ B \end{pmatrix}_{(n+p)\times m} \quad ; \quad A_{n\times m} \boxminus B_{n\times q} := (A|B)_{n\times(m+q)}$$

Seien A_i, B_i , $i=1,..,k$ $(n\times m)$-dimensionale Matrizen:

(1) $\boxplus_i A_i = (\oplus_i A_i) \cdot (1_k \otimes I_m)$

(2) $(\boxplus_i A_i)' = \boxminus_i A_i'$

(3) $\boxplus_i A = 1_k \otimes A$

6. Rechenregeln für M_n

$M_n := (I_{n-1} | -1_{n-1})'$ ist eine $n \times (n-1)$ - dimensionale Matrix

(1) $M_n' \cdot 1_n = 1_n' \cdot M_n = 0_{n-1}$; $M_n' \cdot J_n = 0_{(n-1)\times n}$

(2) $M_n' \cdot M_n = I_{n-1} + J_{n-1}$; $(M_n' \cdot M_n)^{-1} = I_{n-1} - 1/nJ_{n-1}$

(3) $M_n (M_n' M_n)^{-1} M_n' = I_n - 1/nJ_n$

7. Hadamard - Produkt

Seien A,B $(n \times m)$-dimensionale Matrizen:

$$A*B := ((a_{ij} b_{ij}))_{n\times m}$$

In Abschnitt 4.6 wird folgende Matrizenoperation ebenfalls mit dem Zeichen '*' bezeichnet:

Seien $A=(a_1,\ldots,a_k)$, $B=(b_1,\ldots,b_1)$ Matrizen mit n-dimensionalen Spaltenvektoren a_i, b_j. Dann ist

$$A*B := (a_1*b_1, a_1*b_2, .., a_1*b_1, \ldots, a_k*b_1, .., a_k*b_1)$$

eine $(n \times kl)$-dimensionale Matrix mit Hadamard-Produkten der Spaltenvektoren aus A und B.

8. Sonstiges

$J_n \cdot J_n = nJ_n$

A3: EIGENSCHAFTEN VON PERMUTATIONSGRUPPEN

Eine Permutation π vom Grad n, n natürliche Zahl, ist eine eineindeutige (bijektive) lineare Transformation:

$$\pi : R^n \to R^n$$

Die zugehörige Abbildungsmatrix, ebenfalls mit π oder P bezeichnet, ist eine Matrix mit Nullen und Einsen, in der in jeder Zeile und Spalte genau eine Eins vorkommt:

$$P \cdot I_n = I_n \cdot P = P \quad ; \quad P \cdot J_n = J_n = J_n \cdot P \quad ; \quad P \cdot (I_n + aJ_n) = (I_n + aJ_n) \cdot P$$

Sei das kartesische Produkt zweier Permutationsgruppen

$S_k \otimes S_l := \{\pi \in S_{kl} \mid \pi = \pi_k \otimes \pi_l$ und $\pi_k \in S_k, \pi_l \in S_l\}$

Die Anzahl Elemente ist $|S_k \otimes S_l| = k!l!$. $\pi \in S_{kl}$ ist darstellbar als $P_k \otimes P_l$.

Das Kranzprodukt $S_k \wr S_l$ ist darstellbar als $\pi \in S_{kl}$ mit folgenden Eigenschaften:

$$\pi = \begin{pmatrix} M_{11} & \cdots & M_{1l} \\ \cdots & \cdots & \cdots \\ M_{l1} & \cdots & M_{ll} \end{pmatrix}_{kl \times kl}$$

Je Matrix-'Zeile' und Spalte ist genau eine Matrix $M_{ij} \neq 0_{k \times k}$.

Diese ist jeweils eine Permutationsmatrix aus S_l.

Die Anzahl Elemente ist $|S_k \wr S_l| = (k!)^l \cdot l!$.

Sei $f: R^n \to R^n$ eine lineare Transformation; Q_n eine Untergruppe von S_n:

$S_n(f,Q_n) := \{ \pi \in S_n \mid \pi f = f\rho \ , \rho \in Q_n\}$ und

$S_n(f) := \{ \pi \in S_n \mid \pi f = f\rho \ , \rho \in S_n\}$ sind Untergruppen von S_n.

Es gelten folgende Aussagen:

(1) Sei f eine Projektion $S_n(f) = \{\pi \in S_n \mid \pi f = f\pi\}$

(2) $S_n(f) = S_n \Leftrightarrow f \equiv 0$ oder I_n oder $1/nJ_n$ oder E_n

(3) Sei A eine beliebige, quadratische n×n - Matrix:

$S_n(A+aJ_n) = S_n(A)$; $S_n(A+aI_n) = S_n(A)$; $a \in R^1$.

(4) Sei A eine symmetrische, idempotente (Projektions-) Matrix:

$S_n(I_n - A) = S_n(A)$

(5) $A = 1/k J_k \otimes I_l$; $k \cdot l = n$: $S_n(A) = S_k \otimes S_l$

(6) $A=(I_k \otimes 1/lJ_l + 1/kJ_k \otimes I_l)$: $S_n(A) = S_k \otimes S_l$

(7) $A=E_k \otimes E_l = (I_k - 1/kJ_k) \otimes (I_l - 1/lJ_l)$: $S_n(A) = S_k \otimes S_l$

(8) $A=I_l \otimes (I_k - 1/kJ_k)$: $S_n(A) = S_k \sim S_l$

(9) $A=I_l \otimes 1/kJ_k$: $S_n(A) = S_k \sim S_l$.

A4: LINEARE UNTERRÄUME UND PROJEKTIONEN

Sei $U \subset R^n$. U heißt linearer Unterraum des R^n, falls:

(1) mit $x,y \in U$ gilt auch $x+y \in U$ (2) mit x ist auch $-x \in U$

(3) $0_n \in U$.

Seien U_1, U_2 Unterräume des R^n mit $U_1 \cap U_2 = \{0_n\}$ und $\dim(U_1) + \dim(U_2) = n$. Dann ist jedes $x \in R^n$ eindeutig zerlegbar in $x = x_1 + x_2$ mit $x_1 \in U_1$ und $x_2 \in U_2$; d.h. R^n ist darstellbar als direkte Summe von U_1 und U_2 : $U_1 \oplus U_2 = R^n$.

Eine Abbildung f: $x \to x_1$ heißt (orthogonale) Projektion des R^n auf U_1. Diese Abbildung f ist eine lineare Transformation.
Eine lineare Abbildung ist genau dann eine orthogonale Projektion, wenn die zugehörige Abbildungsmatrix K idempotent ist.

Wenn f eine orthogonale Projektion auf U ist, dann ist id - f Projektion des R^n auf das bezüglich R^n orthogonale Komplement von U: $\complement_\perp U$ mit Abbildungsmatrix $I_n - K$. Es gilt $U \oplus \complement_\perp U = R^n$.

Permutationen π: $R^n \to R^n$ sind orthogonale lineare Abbildungen, d.h. $\pi^{-1} = \pi'$.

ANHANG B1

```
*************************************************************************
*                                                                       *
*                    PROGRAM   STROE2                                   *
*                                                                       *
*                                                                       *
*     FORTRAN-VERSION OF THE STREITBERG-ROEHMEL TWO-DIMENSIONAL         *
*     SHIFT ALGORITHM FOR THE TWO-SAMPLE TEST PROBLEM AGAINST           *
*     ONE- OR TWO-SIDED LOCATION ALTERNATIVES                           *
*                                                                       *
*     TEST STATISTIC: SUM OF OBSERVATIONS FOR THE SAMPLE WITH THE       *
*                     SMALLER SAMPLE SIZE                               *
*                     N1<=N2: SAMPLE 1                                  *
*                     N1> N2: SAMPLES ARE EXCHANGED INDICATED BY        *
*                             ISWITCH=1. THE P-VALUES REPORTED HOWEVER  *
*                             REFER TO THE ORIGINAL SAMPLE SIZES        *
*                                                                       *
*                                                                       *
*     INPUT:  1) INTERACTIVE VIA TERMINAL:                              *
*                -INPUT/OUTPUT-FILES, DATA FORMAT (F-FORMAT WITH        *
*                             NO MORE THAN 3 SIGNIFICANT DECIMALS)      *
*                -SAMPLE SIZES N1,N2 ( N1+N2<=100 )                     *
*                -ONE/TWO-SIDED TEST:                                   *
*                             ITAIL=-1 ONE-SIDED FOR H0:  MUE1<=MUE2    *
*                             ITAIL= 1 ONE-SIDED FOR H0:  MUE1>=MUE2    *
*                             ITAIL= 2 TWO-SIDED FOR H0:  MUE1 =MUE2    *
*              2) REAL-TYPE RAW-DATA IN STANDARD RECTANGULAR FORMAT      *
*                                                                       *
*     OUTPUT:     -RAW-DATA AND INTEGER-TRANSFORMED RAW-DATA             *
*                 -VALUE OF TEST STATISTIC FOR 'UNPERMUTED' DATA        *
*                 -PGT: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .GT. ORIGINAL TEST STATISTIC VALUE              *
*                 -PLT: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .LT. ORIGINAL TEST STATISTIC VALUE              *
*                 -PEQ: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .EQ. ORIGINAL TEST STATISTIC VALUE              *
*                 -P-VALUE: ONE/TWO-SIDED P-VALUE                       *
*                       (TWO-SIDED: P-VALUE=MIN(PGT+PEQ,PLT+PEQ,0.5)*2. *
*                 -SP:  FIELD LENGTH REQUIRED FOR GENERATION OF THE     *
*                       'RELEVANT' PART OF THE FREQUENCY-DISTRIBUTION   *
*                       OF TEST STATISTIC VALUES                        *
*                 -CPU_TIME: COMPUTATION TIME REQUIRED FOR CARRYING OUT *
*                       THE SHIFT-ALGORITHM                             *
*                                                                       *
*     REFERENCES: STREITBERG,B.&ROEHMEL,J.(1986)                        *
*                 EXACT DISTRIBUTIONS FOR PERMUTATION AND RANK TESTS:   *
*                 AN INTRODUCTION TO SOME RECENTLY PUBLISHED            *
*                 ALGORITHMS.                                           *
*                 STATISTICAL SOFTWARE NEWSLETTER 12, 10-17.            *
*                                                                       *
*                 KOENIG,J.,GIANI,G.&HAUX,R.(1986)                      *
*                 SOME ALGORITHMS FOR THE DISTRIBUTIONS OF NONPARA-     *
*                 METRIC TWO-SAMPLE TESTS - CONSIDERATIONS ON COM-      *
*                 PLEXITY AND EMPIRICAL INVERSTIGATIONS.                *
*                 PART 1: PERMUTATION TESTS                             *
*                 STATISTICAL SOFTWARE NEWSLETTER 12,20-27.             *
*                                                                       *
*************************************************************************


      INTEGER N1,N2,N,SP,IFEHL,ILEN
      INTEGER CPUBEG,CPUEND
      REAL R(100),PGT,PLT,PEQ,PTWO,PVALUE
      INTEGER IR(100)
      CHARACTER*60 NFMT,FILEIN,FILEOUT,TITLE
      CHARACTER*29 TEXT(0:2)
```

```
C ***************************************************************************
C     INTEGER IR(N1+N2=N)                                                    *
C ***************************************************************************

      DATA TEXT / 'ONE-SIDED  H0: MUE1<=MUE2   :',
     -            'ONE-SIDED  H0: MUE1>=MUE2   :',
     -            'TWO-SIDED  H0: MUE1 =MUE2   :'  /

      CALL CPU_TIM(CPUBEG)

C ***************************************************************************
C     INPUT SECTION                                                          *
C ***************************************************************************

      IN=5
      OUT=6

 20   WRITE(OUT,100)
      READ(IN,101) LENTITLE,TITLE
      IF(LENTITLE.GT.60) GOTO 20
 22   WRITE(OUT,103)
      READ(IN,101) LENFILEIN,FILEIN
      OPEN(UNIT=10,NAME=FILEIN,READONLY,TYPE='OLD',ERR=23)
      GOTO 24
 23   WRITE(OUT,*) '### ERROR OPENING INPUT-FILE ###'
      GOTO 22
 24   WRITE(OUT,104)
      READ(IN,101) LENNFMT,NFMT
      DO 25 I=1,LENNFMT
      IF(NFMT(I:I).EQ.'I'.OR.NFMT(I:I).EQ.'A') THEN
        WRITE(OUT,*) '### INPUT FORMAT MUST BE F-FORMAT ###'
        GOTO 24
      ENDIF
 25   CONTINUE
 26   WRITE(OUT,105)
      READ(IN,101) LENFILEOUT,FILEOUT
      OPEN(UNIT=20,NAME=FILEOUT,TYPE='NEW',ERR=27)
      GOTO 28
 27   WRITE(OUT,*) '### ERROR OPENING OUTPUT-FILE ###'
      GOTO 26
 28   WRITE(OUT,106)
      READ(IN,*) N1,N2

      N = N1 + N2

      IF(N.GT.100) GOTO 28
 29   WRITE(OUT,107)
      READ(IN,*) ITAIL
      IF((ITAIL-1)*(ITAIL+1)*(ITAIL-2).NE.0) GOTO 29

C ***************************************************************************
C     INPUT OF RAW-DATA                                                      *
C ***************************************************************************

      DO I=1,N
        READ(10,NFMT,ERR=33,END=34) R(I)
      END DO
      GOTO 36
33    TYPE 108
      STOP
34    TYPE 109
      STOP

C ***************************************************************************
C     INITIALIZE                                                             *
C ***************************************************************************
```

```
36    SP=0
      DO I=1,N
        IR(I)=0
      END DO

C ***************************************************************************
C     CONVERSION OF REAL-TYPE RAW-DATA TO INTEGER-DATA BY USE OF MUL-   *
C     TIPLICATION FACTOR 10**IFAK. BECAUSE IFAK=3 IN SUBROUTINE         *
C     CONVERT, NO MORE THAN 3 SIGNIFICANT DECIMALS SHOULD BE PRESENT    *
C     IN THE DATA.                                                      *
C ***************************************************************************

      CALL CONVERT(R,IR,N)

C ***************************************************************************
C     ACTUAL PERFORMANCE OF THE TWO-DIMENSIONAL SHIFT ALGORITHM AND     *
C     RETURNING OF P-VALUES FOR ONE- AND TWO-SIDED DECISIONS            *
C ***************************************************************************

      CALL PERMU2(N,N1,N2,IR,ILEN,SP,PGT,PLT,PEQ,PTWO,IFEHL,ISWITCH)

C ***************************************************************************
C     TERMINATION OF PROGRAM, IF SPACE IN ARRAY G( ) OF SUBROUTINE      *
C     TWOSHIFT IS NOT SUFFICIENT TO CARRY OUT THE SHIFT OPERATIONS.     *
C ***************************************************************************

      IF(IFEHL.EQ.1) THEN
        WRITE(OUT,*) '*** MORE SPACE REQUIRED. INCREASE ARRAY G TO ',SP
        WRITE(OUT,*) '*** PROGRAM TERMINATED'
        STOP
      ENDIF

C ***************************************************************************
C     COMPUTATION OF ONE- OR TWO-SIDED P-VALUE                          *
C ***************************************************************************

      IF(ITAIL.EQ.-1) PVALUE=PGT+PEQ
      IF(ITAIL.EQ.1)  PVALUE=PLT+PEQ
      IF(ITAIL.EQ.2)  PVALUE=PTWO

C ***************************************************************************
C     OUTPUT OF RESULTS                                                 *
C ***************************************************************************

      WRITE(20,110) TITLE(1:LENTITLE)

      CALL RAWOUT(NFMT,R,IR,N1,N2,N,FILEIN,ISWITCH)

      WRITE(20,111) ILEN-1
      IF(ITAIL.EQ.-1) ITAIL=0
      WRITE(20,112) PGT,PLT,PEQ,TEXT(ITAIL),PVALUE

      CALL CPU_TIM(CPUEND)
      CPU_TIME=FLOAT(CPUEND-CPUBEG)/100.
      WRITE(20,113) CPU_TIME

      STOP

C---------------------- FORMAT - STATEMENTS ----------------------------

  100 FORMAT(' *** ENTER TITLE (MAX. 60 CHARACTERS)                :',$)
  101 FORMAT(Q,A60)
  103 FORMAT(' *** ENTER NAME OF INPUT FILE                        :',$)
  104 FORMAT(' *** ENTER INPUT FORMAT(MAX. 60 CHARACTERS,F-FORMAT) :',$)
  105 FORMAT(' *** ENTER NAME OF OUTPUT FILE                       :',$)
```

```
  106 FORMAT(' *** ENTER SAMPLE SIZES N1,N2                              :',$)
  107 FORMAT(' *** ENTER -1:   HO : MUE1<=MUE2'/
     -       '            1:   HO : MUE1>=MUE2'/
     -       '            2:   HO : MUE1 =MUE2                          :',$)
  108 FORMAT(//' *** ERROR DETECTED WHEN READING IN RAW DATA ***'/)
  109 FORMAT(//' *** INSUFFICIENT RAW DATA ***'/)
  110 FORMAT(1H1,71(1H*)/1X,1H*,T72,1H*/
     -       1X,1H*,18X,'EXACT TWO SAMPLE PERMUTATION TEST',T72,1H*/
     -       1X,1H*,T72,1H*/1X,1H*,4X,'USING THE STREITBERG-ROEHMEL',
     -       ' TWO-DIMENSIONAL SHIFT ALGORITHM',T72,1H*/1X,1H*,T72,1H*/
     -       1X,1H*,T72,1H*/1X,1H*,<(69-LENTITLE)/2>X,
     -       A<LENTITLE>,T72,1H*/1X,1H*,T72,1H*/1X,71(1H*))
  111 FORMAT(/3X,'VALUE OF TEST STATISTIC (SUM OF FIRST SAMPLE',1H',
     -       'S VALUES) :',I8)
  112 FORMAT(/20X,33(1H*)/20X,'* RESULT OF THE SHIFT ALGORITHM *'/
     -        20X,33(1H*)//
     -        30X,'P> :',F8.4/30X,'P< :',F8.4/30X,'P= :',F8.4//
     -'   NON RANDOMIZED P-VALUE FOR ',A29,F8.4)
  113 FORMAT(/'   REQUIRED CPU-TIME :',F10.2)
C-------------------------------------------------------------------------

      END

C ***************************************************************************

      SUBROUTINE TWOSHIFT(N,NN1,A,G,ILEN,IFEHL)

*****************************************************************************
*                                                                           *
*     SUBROUTINE TWOSHIFT PERFORMS THE ACTUAL COMPUTATION OF THE            *
*     EXACT PERMUTATION DISTRIBUTION                                        *
*     ALGORITHM DESIGNED BY J.KOENIG 1985                                   *
*                                                                           *
*****************************************************************************

      REAL*8 G(ILEN,NN1)
      INTEGER A(200),L(201),R(201)

      IFEHL=0
      N1=NN1-1

C ***************************************************************************
C     CHECK FOR ARRAY-OVERFLOW OF ARRAY G(ILEN,NN1)                         *
C ***************************************************************************
      IF(ILEN*NN1.GT.400000) THEN
        IFEHL=1
        R E T U R N
      END IF

C ***************************************************************************
C     INITIALIZE                                                            *
C ***************************************************************************

      DO 10 I=1,N+1
      L(I)=ILEN
   10 R(I)=0
      DO 20 J=1,N1+1
      DO 30 I=1,ILEN
   30 G(I,J)=0
   20 CONTINUE

C ***************************************************************************
C     SEED FOR THE RECURSION ALGORITHM                                      *
C ***************************************************************************

      G(1,1)=1
```

```
      R(1)=1
      L(1)=1

C ***************************************************************************
C     ACTUAL SHIFT ALGORITHM                                                *
C ***************************************************************************

      DO 40 I=1,N
      IEN1=MIN(I,N1)
      DO 50 K=1,IEN1
      IZENEU=IEN1-K+2
      IZEALT=IZENEU-1
      L(IZENEU)=MIN(L(IZENEU),L(IZEALT)+A(I))
      R(IZENEU)=MAX(R(IZENEU),R(IZEALT)+A(I))
      ILINKS=L(IZEALT)+A(I)
      IRECHT=MIN(R(IZENEU),ILEN)
      DO 60 J=ILINKS,IRECHT
      G(J,IZENEU)=G(J,IZENEU)+G(J-A(I),IZEALT)
   60 CONTINUE
   50 CONTINUE
   40 CONTINUE

      R E T U R N

      END

C ***************************************************************************

      SUBROUTINE PERMU2(N,N1,N2,IR,ILEN,SP,PGT,PLT,PEQ,PTWO,IFEHL,
     -                  ISWITCH)

***************************************************************************
*                                                                         *
*     DRIVER FOR THE COMPUTATION OF THE ONE- AND TWO-SIDED P-VALUES       *
*     FOR THE EXACT TWO SAMPLE PERMUTATION TEST                           *
*                                                                         *
*     THE ACTUAL COMPUTATION OF THE EXACT PERMUTATION DISTRIBUTION        *
*     ( I.E. FREQUENCY OF DIFFERENT VALUES OF THE TEST-STATISTIC )        *
*     IS CARRIED OUT IN SUBROUTINE TWOSHIFT                               *
*                                                                         *
***************************************************************************

      REAL*8 G(400000), FAK *8 ,PUU *8 ,YNOVERM *8,PEE *8
      INTEGER IR(N),B(200),SP

C ***************************************************************************
C     INITIALIZE                                                            *
C ***************************************************************************

      IFEHL=0
      SP=0
      MINA=IR(1)
      ISWITCH=0
      PUU=0

C ***************************************************************************
C     FIND MINIMUM OF TRANSFORMED INTEGER DATA                              *
C ***************************************************************************

      DO I=1,N
        IF(IR(I).LT.MINA) MINA=IR(I)
      END DO

C ***************************************************************************
C     TRANSFORM DATA TO NON-NEGATIVE INTEGER DATA                           *
C ***************************************************************************
```

```
      DO I=1,N
        B(I)=IR(I)-MINA
      END DO

C *************************************************************************
C     EXCHANGE SAMPLE 1 AND SAMPLE 2 , IF N1 > N2                         *
C *************************************************************************

      IF(N1.GT.N-N1)THEN
        ISWITCH=1
        N1=N-N1
        N2=N-N1
        JAN=N/2
        DO IJ=1,JAN
          JAP=B(IJ)
          B(IJ)=B(N-IJ+1)
          B(N-IJ+1)=JAP
        END DO
      ENDIF

C *************************************************************************
C     COMPUTE NUMBER OF ELEMENTS SP IN TWO-DIM. ARRAY G(N1+1,ILEN+1)      *
C     FOR THE TWO-DIMENSIONAL SHIFT ALGORITHM                             *
C     N1+1                    ROWS                                        *
C     ILEN=1+B(1)+...+B(N1) COLUMNS                                       *
C *************************************************************************

      ILEN=1

      DO I=1,N1
        ILEN=ILEN+B(I)
      END DO

      N1P=N1+1
      SP=ILEN*N1P

C *************************************************************************
C     ACTUAL COMPUTATION OF THE EXACT PERMUTATION DISTRIBUTION            *
C *************************************************************************

      CALL TWOSHIFT (N,N1P,B,G,ILEN,IFEHL)

      IF(IFEHL.EQ.1) RETURN

C *************************************************************************
C     EXACT PROBABILITY FOR EACH ADMISSIBLE PERMUTATION OF THE DATA       *
C *************************************************************************

      FAK=YNOVERM(N,N1)
      GO=ILEN*N1

C *************************************************************************
C     USE THE ROW OF G CORRESPONDING TO THE SAMPLE SIZE OF N1             *
C *************************************************************************

      DO 50 I=GO+1,GO+ILEN-1
   50 PUU=PUU+G(I)
      PEE=G(GO+ILEN)

C *************************************************************************
C     COMPUTATION OF P-VALUES                                             *
C *************************************************************************

      PLT  = FAK*PUU
      PEQ  = FAK*PEE
```

```
      PGT  = 1.-PLT-PEQ
      PTWO = MIN(PLT+PEQ,PGT+PEQ)*2.
      PTWO = MIN(1.,PTWO)
      IF(ISWITCH.EQ.1) THEN
         PGT = PLT
         PLT = 1.-PGT-PEQ
      END IF

      R E T U R N

      END
C ***************************************************************************

      DOUBLE PRECISION FUNCTION YNOVERM(N,M)

*****************************************************************************
*                                                                           *
*     DOUBLE PRECISION FUNCTION YNOVERM COMPUTES N OVER M                   *
*                                                                           *
*****************************************************************************

      REAL*8 FAK

      FAK=0

      DO I=1,M
        FAK=FAK+DLOG(DFLOAT(N-I+1))-DLOG(DFLOAT(M-I+1))
      END DO

      FAK=-FAK

      YNOVERM=DEXP(FAK)

      R E T U R N

      END
C ***************************************************************************

      SUBROUTINE CONVERT ( R , IR , N )

*****************************************************************************
*                                                                           *
*     SUBROUTINE CONVERT TRANSFORMS REAL NUMBERS FROM ARRAY R(N)            *
*     WITH NO MORE THAN 3 DECIMALS INTO INTEGER NUMBERS IN ARRAY IR(N)      *
*                                                                           *
*****************************************************************************

      INTEGER N,IR(N)
      REAL R(N)

      DO I=1,N
        IR(I)=NINT(R(I)*1000)
      END DO

C *****************************************************************************
C     IF MORE THAN THREE DECIMALS ARE PRESENT IFAK MUST BE INCREASED          *
C *****************************************************************************

      IFAK=3

    5 K=0
      DO I=1,N
        IF(MOD(IR(I),10).EQ.0)        K=K+1
      END DO
```

```
      IF(K.EQ.N) THEN
         DO I=1,N
           IR(I)=NINT(IR(I)/10.)
         END DO
         IFAK=IFAK-1
         GOTO 5
      END IF
      IF(K.NE.N)   R E T U R N

      END

C *************************************************************************

      SUBROUTINE RAWOUT(NFMT,Y,IY,N1,N2,NSUM,FILEIN,ISWITCH)

C *************************************************************************
C     IN SUBROUTINE RAWOUT THE RAW DATA OF AN INDEPENDENT TWO-SAMPLE      *
C     PROBLEM AREOUTPUT TO LOGICAL UNIT 20                                *
C *************************************************************************

      INTEGER N(2),NSUM,IY(100)
      REAL Y(100)

      CHARACTER*60 FILEIN,NFMT

      IF(ISWITCH.EQ.0) THEN
        N(1)=N1
        N(2)=N2
      ELSE
        N(1)=N2
        N(2)=N1
      END IF

      WRITE(20,100) FILEIN,NFMT

      WRITE(20,104)

      NZAEHL=0
      DO L=1,2
        DO I=1,N(L)
          NZAEHL=NZAEHL+1
          IF(I.NE.(N(L)+1)/2) THEN
            WRITE(20,102) I,Y(NZAEHL)
          ELSE
            WRITE(20,103) L,I,Y(NZAEHL)
          END IF
        END DO
        WRITE(20,104)
      END DO

      WRITE(20,105)

      WRITE(20,104)

      NZAEHL=0
      DO L=1,2
        DO I=1,N(L)
          NZAEHL=NZAEHL+1
          IF(I.NE.(N(L)+1)/2) THEN
            WRITE(20,106) I,IY(NZAEHL)
          ELSE
            WRITE(20,107) L,I,IY(NZAEHL)
          END IF
        END DO
        WRITE(20,104)
      END DO
```

```
      R E T U R N

C ---------------------------- FORMAT - STATEMENTS ---------------------

  100 FORMAT(/25X,21(1H*)/25X,1H*,' R A W  -  D A T A ',1H*/25X,
     221(1H*)//12X,'INPUT FILE  : ',A60/12X,'DATA-FORMAT : ',A60/)
  102 FORMAT(20X,'I',10X,'I',I3,' I',2X,F10.3,'  I')
  103 FORMAT(20X,'I SAMPLE ',I1,' I',I3,' I',2X,F10.3,'  I')
  104 FORMAT(20X,'I',18(1H-),(10(1H-)),'--I')
  105 FORMAT(/20X,32(1H*)/20X,'* INTEGER TRANSFORMED RAW-DATA *'/
     -        20X,32(1H*)/)
  106 FORMAT(20X,'I',10X,'I',I3,' I',2X,I10,'  I')
  107 FORMAT(20X,'I SAMPLE ',I1,' I',I3,' I',2X,I10,'  I')
C ---------------------------------------------------------------------

      END
```

A N H A N G B2

```
*************************************************************************
*                                                                       *
*                     PROGRAM   STROE1                                  *
*                                                                       *
*                                                                       *
*                                                                       *
*     FORTRAN-VERSION OF THE STREITBERG-ROEHMEL ONE-DIMENSIONAL         *
*     SHIFT ALGORITHM FOR THE DEPENDENT TWO-SAMPLE TEST PROBLEM         *
*     AGAINST ONE- OR TWO-SIDED LOCATION ALTERNATIVES                   *
*                                                                       *
*     TEST STATISTIC: SUM OF FIRST OBSERVATIONS ACROSS ALL BIVARIATE    *
*                     SAMPLE ELEMENTS                                   *
*                                                                       *
*                                                                       *
*     INPUT:  1) INTERACTIVE VIA TERMINAL:                              *
*               -INPUT/OUTPUT-FILES, DATA FORMAT (F-FORMAT WITH         *
*                          NO MORE THAN 3 SIGNIFICANT DECIMALS)         *
*               -SAMPLE SIZES N ( N<=50 )                               *
*               -ONE/TWO-SIDED TEST:                                    *
*                          ITAIL=-1 ONE-SIDED AGAINST  MUE1<=MUE2       *
*                          ITAIL= 1 ONE-SIDED AGAINST  MUE1>=MUE2       *
*                          ITAIL= 2 TWO-SIDED AGAINST  MUE1<>MUE2       *
*             2) REAL-TYPE RAW-DATA IN STANDARD RECTANGULAR FORMAT      *
*                                                                       *
*                                                                       *
*     OUTPUT:     -RAW-DATA AND INTEGER-TRANSFORMED RAW-DATA            *
*                 -VALUE OF TEST STATISTIC FOR 'UNPERMUTED' DATA        *
*                 -PGT: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .GT. ORIGINAL TEST STATISTIC VALUE              *
*                 -PLT: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .LT. ORIGINAL TEST STATISTIC VALUE              *
*                 -PEQ: RELATIVE FREQUENCY OF TEST STATISTIC VALUES     *
*                       .EQ. ORIGINAL TEST STATISTIC VALUE              *
*                 -P-VALUE: ONE/TWO-SIDED P-VALUE                       *
*                       (TWO-SIDED: P-VALUE=MIN(PGT+PEQ,PLT+PEQ,0.5)*2. *
*                 -SP:  FIELD LENGTH REQUIRED FOR GENERATION OF THE     *
*                       'RELEVANT' PART OF THE FREQUENCY-DISTRIBUTION   *
*                       OF TEST STATISTIC VALUES                        *
*                 -CPU_TIME: COMPUTATION TIME REQUIRED FOR CARRYING OUT *
*                       THE SHIFT-ALGORITHM                             *
*                                                                       *
*                                                                       *
*     REFERENCES: STREITBERG,B.&ROEHMEL,J.(1986)                        *
*                 EXACT DISTRIBUTIONS FOR PERMUTATION AND RANK TESTS:   *
*                 AN INTRODUCTION TO SOME RECENTLY PUBLISHED            *
*                 ALGORITHMS.                                           *
*                 STATISTICAL SOFTWARE NEWSLETTER 12,10-17.             *
*                                                                       *
*                 KOENIG,J.,GIANI,G.&HAUX,R.(1986)                      *
*                 SOME ALGORITHMS FOR THE DISTRIBUTIONS OF NONPARA-     *
*                 METRIC TWO-SAMPLE TESTS - CONSIDERATIONS ON COM-      *
*                 PLEXITY AND EMPIRICAL INVERSTIGATIONS.                *
*                 PART 1: PERMUTATION TESTS                             *
*                 STATISTICAL SOFTWARE NEWSLETTER 12,20-27.             *
*                                                                       *
*************************************************************************

      INTEGER N,SP,IFEHL,ILEN,GO
      INTEGER CPUBEG,CPUEND
      INTEGER IR(100)
      REAL R(100),PGT,PLT,PEQ,PTWO,PVALUE
      CHARACTER*60 NFMT,FILEIN,FILEOUT,TITLE
      CHARACTER*29 TEXT(0:2)

C *************************************************************************
C     INTEGER IR(2*N)
```

```
C ***************************************************************************

      DATA TEXT / 'ONE-SIDED TEST MUE1<=MUE2   :',
     -            'ONE-SIDED TEST MUE1>=MUE2   :',
     -            'TWO-SIDED TEST MUE1<>MUE2   :'  /

      CALL CPU_TIM(CPUBEG)

C ***************************************************************************
C     INPUT SECTION                                                        *
C ***************************************************************************

      IN=5
      OUT=6

 20   WRITE(OUT,100)
      READ(IN,101) LENTITLE,TITLE
      IF(LENTITLE.GT.60) GOTO 20
 22   WRITE(OUT,103)
      READ(IN,101) LENFILEIN,FILEIN
      OPEN(UNIT=10,NAME=FILEIN,READONLY,TYPE='OLD',ERR=23)
      GOTO 24
 23   WRITE(OUT,*) '### ERROR OPENING INPUT-FILE ###'
      GOTO 22
 24   WRITE(OUT,104)
      READ(IN,101) LENNFMT,NFMT
      DO 25 I=1,LENNFMT
      IF(NFMT(I:I).EQ.'I'.OR.NFMT(I:I).EQ.'A') THEN
        WRITE(OUT,*) '### INPUT FORMAT MUST BE F-FORMAT ###'
        GOTO 24
      ENDIF
 25   CONTINUE
 26   WRITE(OUT,105)
      READ(IN,101) LENFILEOUT,FILEOUT
      OPEN(UNIT=20,NAME=FILEOUT,TYPE='NEW',ERR=27)
      GOTO 28
 27   WRITE(OUT,*) '### ERROR OPENING OUTPUT-FILE ###'
      GOTO 26
 28   WRITE(OUT,106)
      READ(IN,*) N
      IF(N.GT.50) GOTO 28
 29   WRITE(OUT,107)
      READ(IN,*) ITAIL
      IF((ITAIL-1)*(ITAIL+1)*(ITAIL-2).NE.0) GOTO 29

C ***************************************************************************
C     INPUT OF RAW-DATA                                                    *
C ***************************************************************************

      DO I=1,N
        READ(10,NFMT,ERR=33,END=34) R(I),R(N+I)
      END DO
      GOTO 36
33    TYPE 108
      STOP
34    TYPE 109
      STOP

C ***************************************************************************
C     INITIALIZE                                                           *
C ***************************************************************************

36    DO I=1,N
        IR(I)=0
        IR(N+I)=0
      END DO
```

```
C ************************************************************************
C     CONVERSION OF REAL-TYPE RAW-DATA TO INTEGER-DATA USING MULTI-    *
C     PLICATION FACTOR 10**IFAK. BECAUSE IFAK=3 IN SUBROUTINE CONVERT *
C     ,NO MORE THAN 3 SIGNIFICANT DECIMALS SHOULD BE PRESENT IN THE    *
C     DATA.                                                            *
C ************************************************************************

      CALL CONVERT(R,IR,N)

C ************************************************************************
C     ACTUAL PERFORMANCE OF THE ONE-DIMENSIONAL SHIFT ALGORITHM AND    *
C     RETURNING OF P-VALUES FOR ONE- AND TWO-SIDED DECISIONS           *
C ************************************************************************

      CALL PERMU1(N,IR,ILEN,GO,SP,PGT,PLT,PEQ,PTWO,IFEHL)

C ************************************************************************
C     TERMINATION OF PROGRAM, IF SPACE IN VECTOR G( ) OF SUBROUTINE    *
C     ONESHIFT IS NOT SUFFICIENT TO CARRY OUT THE SHIFT OPERATIONS     *
C ************************************************************************

      IF(IFEHL.EQ.1) THEN
        WRITE(OUT,*) '** MORE SPACE REQUIRED. INCREASE VECTOR G TO ',SP
        WRITE(OUT,*) '** PROGRAM TERMINATED'
        STOP
      ENDIF

C ************************************************************************
C     COMPUTATION OF ØNE- OR TWO-SIDED P-VALUE                         *
C ************************************************************************

      IF(ITAIL.EQ.-1) PVALUE=PLT+PEQ
      IF(ITAIL.EQ.1)  PVALUE=PGT+PEQ
      IF(ITAIL.EQ.2)  PVALUE=PTWO

C ************************************************************************
C     OUTPUT OF RESULTS                                                *
C ************************************************************************

      WRITE(20,110) TITLE(1:LENTITLE)
      CALL RAWOUT(NFMT,R,IR,N,FILEIN)
      WRITE(20,111) GO
      IF(ITAIL.EQ.-1) ITAIL=0
      WRITE(20,112) PLT,PGT,PEQ,TEXT(ITAIL),PVALUE

      CALL CPU_TIM(CPUEND)
      CPU_TIME=FLOAT(CPUEND-CPUBEG)/100.
      WRITE(20,113) CPU_TIME
      STOP

C---------------------- FORMAT - STATEMENTS ----------------------------

  100 FORMAT(' *** ENTER TITLE (MAX. 60 CHARACTERS)                :',$)
  101 FORMAT(Q,A60)
  103 FORMAT(' *** ENTER NAME OF INPUT FILE                        :',$)
  104 FORMAT(' *** ENTER INPUT FORMAT(MAX. 60 CHARACTERS,F-FORMAT) :',$)
  105 FORMAT(' *** ENTER NAME OF OUTPUT FILE                       :',$)
  106 FORMAT(' *** ENTER SAMPLE SIZES N                            :',$)
  107 FORMAT(' *** ENTER -1: MUE1<=MUE2  1:MUE1>=MUE2  2:MUE1<>MUE2:',$)
  108 FORMAT(//' *** ERROR DETECTED WHEN READING IN RAW DATA ***'/)
  109 FORMAT(//' *** INSUFFICIENT RAW DATA ***'/)
  110 FORMAT(1H1,71(1H*)/1X,1H*,T72,1H*/
     -       1X,1H*,18X,'EXACT TWO SAMPLE PERMUTATION TEST',T72,1H*/
     -       1X,1H*,T72,1H*/1X,1H*,4X,'USING THE STREITBERG-ROEHMEL',
     -       ' ONE-DIMENSIONAL SHIFT ALGORITHM',T72,1H*/1X,1H*,T72,1H*/
```

```
     -       1X,1H*,T72,1H*/1X,1H*,<(69-LENTITLE)/2>X,
     -       A<LENTITLE>,T72,1H*/1X,1H*,T72,1H*/1X,71(1H*))
  111 FORMAT(/3X,'VALUE OF TEST STATISTIC (SUM OF FIRST SAMPLE',1H',
     -       'S VALUES) :',I8)
  112 FORMAT(/20X,33(1H*)/20X,'* RESULT OF THE SHIFT ALGORITHM *'/
     -        20X,33(1H*)//
     -        30X,'P< :',F8.4/30X,'P> :',F8.4/30X,'P= :',F8.4//
     -'   NON RANDOMIZED P-VALUE FOR ',A29,F8.4)
  113 FORMAT(/'   REQUIRED CPU-TIME :',F10.2)
C-----------------------------------------------------------------------

      END

C ***********************************************************************

      SUBROUTINE ONESHIFT(N,B,G,ILEN,IFEHL)

************************************************************************
*                                                                      *
*     SUBROUTINE ONESHIFT PERFORMS THE ACTUAL COMPUTATION OF THE       *
*     EXACT PERMUTATION DISTRIBUTION                                   *
*                                                                      *
*     ALGORITHM DESIGNED BY H.RATAJCZAK 1986                           *
*                                                                      *
************************************************************************

      REAL*8 G(2*ILEN)
      INTEGER B(200)

      IFEHL=0

C ***********************************************************************
C     CHECK FOR ARRAY-OVERFLOW                                         *
C ***********************************************************************

      IF(ILEN*2.GT.400000) THEN
        IFEHL=1
        R E T U R N
      END IF

C ***********************************************************************
C     ACTUAL SHIFT ALGORITHM                                           *
C ***********************************************************************

C ***********************************************************************
C     SEED FOR THE SHIFT ALGORITHM                                     *
C ***********************************************************************

      G(1)=1
      NCOL=1

      DO 40 I=1,N
        DO 10 K=1,B(I)
        G(NCOL+K)=0
10      G(NCOL+B(I)+K)=0

C ***********************************************************************
C     SHIFT OPERATION                                                  *
C ***********************************************************************

        DO 20 K=1,NCOL
20      G(NCOL+2*B(I)+K)=G(K)
        NCOL=NCOL+B(I)
C ***********************************************************************
C     ADDITION                                                         *
C ***********************************************************************
```

```
        DO 30 K=1,NCOL
30      G(K)=G(K)+G(NCOL+K)
40    CONTINUE

      R E T U R N

      END

C *************************************************************************

      SUBROUTINE PERMU1(N,IR,ILEN,GO,SP,PGT,PLT,PEQ,PTWO,IFEHL)

**************************************************************************
*                                                                        *
*     DRIVER FOR THE COMPUTATION OF THE ONE- AND TWO-SIDED P-VALUES      *
*     FOR THE EXACT TWO SAMPLE PERMUTATION TEST                          *
*                                                                        *
*     THE ACTUAL COMPUTATION OF THE EXACT PERMUTATION DISTRIBUTION       *
*     ( ??? FREQUENCY OF DIFFERENT VALUES OF THE TEST-STATISTIC )        *
*     IS CARRIED OUT IN SUBROUTINE ONESHIFT                              *
*                                                                        *
**************************************************************************

      REAL*8 G(400000), FAK *8 ,PUU *8 ,PEE *8
      INTEGER IR(2*N),B(200),SP,GO

C *************************************************************************
C     INITIALIZE                                                          *
C *************************************************************************

      GO=0
      PUU=0
      ILEN=1

C *************************************************************************
C     FIND MINIMUM OF EACH PAIR OF TRANSFORMED INTEGER DATA               *
C *************************************************************************

      DO I=1,N
        MINA=MIN(IR(I),IR(N+I))
        GO=GO+IR(I)-MINA
        B(I)=MAX(IR(I)-MINA,IR(N+I)-MINA)
        ILEN=ILEN+B(I)
      END DO

      SP=ILEN*2
C *************************************************************************
C     ACTUAL COMPUTATION OF THE EXACT PERMUTATION DISTRIBUTION            *
C *************************************************************************

      CALL ONESHIFT (N,B,G,ILEN,IFEHL)

      IF(IFEHL.EQ.1) RETURN

C *************************************************************************
C     EXACT PROBABILITY FOR EACH ADMISSIBLE PERMUTATION OF THE DATA       *
C *************************************************************************

      FAK=1./2.**FLOAT(N)

      DO 50 I=GO+2,ILEN
   50 PUU=PUU+G(I)
      PEE=G(GO+1)

C *************************************************************************
```

```
C     COMPUTATION OF P-VALUES                                           *
C ***************************************************************************

      PLT  = FAK*PUU
      PEQ  = FAK*PEE
      PGT  = 1.-PLT-PEQ
      PTWO = MIN(PLT+PEQ,PGT+PEQ)*2.
      PTWO = MIN(1.,PTWO)

      R E T U R N

      END

C ***************************************************************************

      SUBROUTINE CONVERT ( R , IR , N )

*****************************************************************************
*                                                                           *
*     SUBROUTINE CONVERT TRANSFORMS REAL NUMBERS FROM ARRAY R(N)            *
*     WITH NO MORE THAN 3 DECIMALS INTO INTEGER NUMBERS IN ARRAY IR(N)      *
*                                                                           *
*****************************************************************************

      INTEGER N,IR(2*N)
      REAL R(2*N)

      DO I=1,2*N
        IR(I)  =NINT(R(I)*1000)
      END DO

C ***************************************************************************
C     INCREASE IFAK IF MORE THAN THREE DECIMALS ARE PRESENT                 *
C ***************************************************************************

      IFAK=3

    5 K=0
      DO I=1,2*N
        IF(MOD(IR(I),10).EQ.0) K=K+1
      END DO
      IF(K.EQ.2*N) THEN
         DO I=1,2*N
           IR(I)=NINT(IR(I)/10.)
         END DO
         IFAK=IFAK-1
         GOTO 5
      END IF

      IF(K.NE.2*N)   R E T U R N

      END

C ***************************************************************************

      SUBROUTINE RAWOUT(NFMT,Y,IY,NSUM,FILEIN)

C ***************************************************************************
C     IN SUBROUTINE RAWOUT THE RAW DATA OF A DEPENDENT TWO-SAMPLE DESIGN    *
C     ARE OUTPUT TO LOGICAL UNIT 20                                         *
C ***************************************************************************

      INTEGER NSUM,IY(100),N(2)
      REAL Y(100)

      CHARACTER*60 FILEIN,NFMT
```

```
      N(1)=NSUM
      N(2)=NSUM

      WRITE(20,100) FILEIN,NFMT

      WRITE(20,104)

      NZAEHL=0
      DO L=1,2
        DO I=1,N(L)
          NZAEHL=NZAEHL+1
          IF(I.NE.(N(L)+1)/2) THEN
            WRITE(20,102) I,Y(NZAEHL)
          ELSE
            WRITE(20,103) L,I,Y(NZAEHL)
          END IF
        END DO
        WRITE(20,104)
      END DO

      WRITE(20,105)
      WRITE(20,104)

      NZAEHL=0
      DO L=1,2
        DO I=1,N(L)
          NZAEHL=NZAEHL+1
          IF(I.NE.(N(L)+1)/2) THEN
            WRITE(20,106) I,IY(NZAEHL)
          ELSE
            WRITE(20,107) L,I,IY(NZAEHL)
          END IF
        END DO
        WRITE(20,104)
      END DO

      R E T U R N

C ---------------------------- FORMAT - STATEMENTS ---------------------

  100 FORMAT(/25X,21(1H*)/25X,1H*,' R A W  -  D A T A ',1H*/25X,
     221(1H*)//12X,'INPUT FILE  : ',A60/12X,'DATA-FORMAT : ',A60/)
  102 FORMAT(20X,'I',10X,'I',I3,' I',2X,F10.3,'  I')
  103 FORMAT(20X,'I SAMPLE ',I1,' I',I3,' I',2X,F10.3,'  I')
  104 FORMAT(20X,'I',18(1H-),(10(1H-)),'--I')
  105 FORMAT(/20X,32(1H*)/20X,'* INTEGER TRANSFORMED RAW-DATA *'/
     -        20X,32(1H*)/)
  106 FORMAT(20X,'I',10X,'I',I3,' I',2X,I10,'  I')
  107 FORMAT(20X,'I SAMPLE ',I1,' I',I3,' I',2X,I10,'  I')
C ---------------------------------------------------------------------

      END
```

ANHANG B3

```
************************************************************************
*                                                                      *
************************************************************************

      SUBROUTINE SEQMTP(KD,EPS,NPERM,IQUOT,INDEX,VERTEILS,PERMNUM,IPGR,
     1IPEQ,ALPHA)

************************************************************************
*                                                                      *
*    SUBROUTINE SEQMTP PERFORMS A SEQUENTIAL MULTIPLE TEST PROCEDURE   *
*    FOR MULTIVARIATE PERMUTATION TESTS USING THE UNIVARIATE P-VALUES  *
*    FOR SETTING UP THE SEQUENCE OF SEQUENTIAL TESTS.                  *
*                                                                      *
*    ALGORITHM DESIGNED BY K.WILLMES AND PROGRAMMED BY H.RATAJCZAK     *
*                                                        /07/85        *
*                                                                      *
************************************************************************

      INTEGER IQUOT(10),INDEX(10),PERMNUM(10,20000),IPGR(10),IPEQ(10)
      INTEGER HILF(20001)
      REAL VERTEILS(10,20000)

************************************************************************
*                                                                      *
*          M A X I M U M   P A R A M E T E R   V A L U E S             *
*                                                                      *
*    KD                  : 10                                          *
*                                                                      *
*    NPERM               : 20000  (19999 PERMUTATIONS OF OBSERVATIONS  *
*                                  PLUS THE ORIGINAL OSERVATIONS FOR   *
*                                  THE MONTE CARLO SOLUTION)           *
*                                                                      *
************************************************************************
*                                                                      *
*          D E S C R I P T I O N   O F   P A R A M E T E R S           *
*                                                                      *
*    ALPHA               : NOMINAL MULTIPLE TYPE-I ERROR LEVEL         *
*                                                                      *
*    NPERM               : MAXIMUM NUMBER OF RANDOM OR SYSTEMATIC PER- *
*                          MUTATIONS GENERATED FOR THE MONTE-CARLO OR  *
*                          THE EXACT SOLUTION OF THE SEQUENTIAL        *
*                          MULTIPLE TEST PROCEDURE                     *
*                          ( NPERM <= 19999 )                          *
*                                                                      *
*    EPS                 : MAXIMUM VALUE OF DIFFERENCE BETWEEN TEST    *
*                          STATISTIC VALUES TAKEN TO BE ZERO           *
*                          (E.G. 1.0E-4)                               *
*                                                                      *
*    KD                  : NUMBER OF DEPENDENT VARIABLES               *
*                          (2 <= KD <= 10)                             *
*                                                                      *
*    IPGR(KD)            : NUMBER OF PERMUTED TEST STATISTIC VALUES    *
*                          FOR DEPENDENT VARIABLE I=1...KD GREATER     *
*                          THAN THE ORIGINAL TEST STATISTIC VALUE      *
*                                                                      *
*    IPEQ(KD)            : NUMBER OF PERMUTED TEST STATISTIC VALUES    *
*                          FOR DEPENDENT VARIABLE I=1...KD EQUAL TO    *
*                          THE ORIGINAL TEST STATISTIC VALUE           *
*                                                                      *
*    INDEX(KD)           : ORIGINAL DEPENDENT VARIABLE SEQUENCE NUMBER *
*                          CORRESPONDING TO THE SMALLEST,SECOND SMALL- *
*                          EST,..,KD-SMALLEST P-VALUE FOR THE KD UNI-  *
*                          VARIATE PERMUTATION TESTS                   *
*                                                                      *
*    VERTEILS(KD,NPERM): ORIGINAL AND "PERMUTED" TEST STATISTIC        *
*                          VALUES ORDERED FROM LOWEST TO HIGHEST FOR   *
```

```
*                          EACH DEPENDENT VARIABLE 1...KD SEPARATELY     *
*                                                                        *
*    PERMNUM(KD,NPERM) : ORDER-OF-GENERATION SEQUENCE NUMBER OF THE      *
*                        KD-VECTORPERMUTATIONS CORRESPONDING TO THE      *
*                        ORDERED TEST STATISTIC VALUES FOR EACH DE-      *
*                        PENDENT VARIABLE VARIABLE 1...KD                *
*                                                                        *
*** IQUOT(KD) ***      : NUMBER OF PERMUTATIONS WHICH JOINTLY(!)         *
*                        FULFILL FOR ORDERED VARIABLES                   *
*                        INDEX(I) <= INDEX(J) <= INDEX(KD), I=1..KD:     *
*                        THE VALUE OF THE UNIVARIATE P-VALUE FUNCTION    *
*                        FOR THE ORIGINAL OBSERVATIONS PERMUTED AC-      *
*                        CORDING TO THAT PERMUTATION ARE JOINTLY,        *
*                        I.E. FOR ALL ORDERED VARIABLES INDEX(I)...      *
*                        INDEX(KD) LESS THAN OR EQUAL TO THE UNIVA-      *
*                        RIATE P-VALUE FOR THE VARIABLE INDEX(I).        *
*                        IF IQUOT(K) IS DIVIDED BY THE NUMBER OF PER-    *
*                        MUTATIONS CARRIED OUT IN THE CALLING PRO-       *
*                        GRAM, IT GIVES THE MULTIPLE P-VALUE FOR THE     *
*                        K-TH STEP OF THE SEQUENTIAL MULTIPLE TEST       *
*                        PROCEDURE K=1...KD.                             *
*                                                                        *
**************************************************************************


    DO 7 MMM=1,KD-1

**************************************************************************
*                                                                        *
*   THE SEQUENTIAL MULTIPLE TEST PROCEDURE STARTS WITH THE DIMENSION     *
*   RELATED TO THE SMALLEST P-VALUE AMONG THE DEPENDENT VARIABLES        *
*   THAT ARE LEFT IN THE SEQUENCE.                                       *
*   NOTE: THE LAST,KD-TH MULTIPLE P-VALUE IS IDENTICAL TO THE SIMPLE     *
*         LARGEST P-VALUE FOR VARIABLE INDEX(KD).                        *
*                                                                        *
**************************************************************************

    I = INDEX(MMM)

**************************************************************************
*   IP IS THE NUMBER OF PERMUTATIONS WHICH LED TO TEST STATISTIC         *
*   VALUES FOR PERMUTED DATA WHICH WERE >= THE INITIAL TEST STATISTIC    *
*   VALUE FOR THE MMM-TH SMALLEST UNIVARIATE P-VALUE.                    *
*                                                                        *
**************************************************************************

    IP = IPGR(I) + IPEQ(I)

    IF( IP .EQ. NPERM ) THEN
          IF(ALPHA.EQ.1.000) THEN
               DO 77 JJJ=MMM,KD-1
 77            IQUOT(JJJ)=NPERM
              GOTO 88
          ELSE

**************************************************************************
*   SET ALL MULTIPLE P-VALUES TO BE >= THE NOMINAL MULTIPLE TYPE-I       *
*   ERROR LEVEL   (N.S.)                                                 *
**************************************************************************

               DO 78 JJJ=MMM,KD
 78            IQUOT(JJJ)=-99
               RETURN
            END IF
    END IF
```

```
************************************************************************
*   INITIALIZE COUNT FOR MMM-TH MULTIPLE P-VALUE                       *
************************************************************************

      NANZ = 0

************************************************************************
*   START COMPUTATIONS FOR THE MMM-TH MULTIPLE P-VALUE                 *
************************************************************************

      DO 1 M=NPERM,NPERM-IP+1,-1
      NANZ = NANZ + 1

************************************************************************
*   STORE INDICES OF PERMUTATIONS LEADING TO TEST STATISTIC VALUES     *
*   >= INITIAL VALUES FOR THE DEPENDENT VARIABLES RELATED TO THE       *
*   SMALLEST P-VALUE AMONG THE DEPENDENT VARIABLES LEFT IN THE         *
*   SEQUENTIAL TEST PROCEDURE (I.E. RELATED TO THE MMM-TH SMALLEST     *
*   P-VALUE).                                                          *
************************************************************************

    1 HILF(NANZ) = PERMNUM(I,M)

************************************************************************
*   DETERMINE THE PERMUTATIONS WHICH JOINTLY LEAD TO P-VALUES <= THE   *
*   MMM-TH SMALLEST P-VALUE FOR ALL DEPENDENT VARIABLES LOWER IN THE   *
*   SEQUENCE.                                                          *
************************************************************************

      DO 6 LLL = MMM+1,KD
      K = INDEX(LLL)
      NP = NPERM - IP + 1

************************************************************************
*   CHECK WHETHER THE P-LEVEL OF THE MMM-TH DEPENDENT VARIABLE CAN BE  *
*   ATTAINED BY DEPENDENT VARIABLES LOWER IN THE HIERARCHY THAN THAT   *
*   VARIABLE.                                                          *
************************************************************************

      IF(VERTEILS(K,NP-1).LT.(VERTEILS(K,NP)-EPS)) GOTO 3

************************************************************************
*   NP INDICATES THE P-LEVEL THAT IS ATTAINABLE FOR DIMENSION LLL.     *
************************************************************************

    2 NP = NP + 1

************************************************************************
*   IF THE P-LEVEL OF THE MMM-TH DEPENDENT VARIABLE IS NOT ATTAINABLE  *
*   BY DIMENSION LLL SKIP THAT DEPENDENT VARIABLE.                     *
************************************************************************

      IF(NP.GT.NPERM) GOTO 6

************************************************************************
*   CHECK WHETHER THE REDUCED P-LEVEL IS ATTAINABLE FOR VARIABLE LLL.  *
************************************************************************

      IF(VERTEILS(K,NP-1).LT.(VERTEILS(K,NP)-EPS)) GOTO 3

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!! WARNING: IF IP=NPERM I.E. THE MOST 'INSIGNIFICANT' P-LEVEL      !!!
!!!          P=1.0, THEN NP-1=0. THIS IS VERY UNLIKELY; PROVISION  !!!
!!!          COULD BE MADE AT THE BEGINNING OF THE SUBROUTINE.      !!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

```
      GOTO 2

    3 DO 5 M=NP,NPERM
      NANZ = NANZ + 1
      HILF(NANZ) = PERMNUM(K,M)

*************************************************************************
*   CHECK WHETHER PERMUTATION INDEX HILF(NANZ) IS ALREADY AMONG THE     *
*   PERMUTATION-INDICES RETAINED SO FAR.                                *
*                                                                       *
*   NOTE: THIS CHECK SLOWS DOWN THE ALGORITHM. BUT IT PREVENTS NANZ     *
*   FROM BECOMING LARGER THAN NPMAX WHICH WOULD CAUSE AN OVERFLOW       *
*   IN ARRAY HILF().                                                    *
*   IF THE MULTIPLE P-VALUES, HOWEVER, ARE COMPARED TO THE MULTIPLE     *
*   ALPHA-LEVEL OF THE TEST PROCEDURE AND THE MULTIPLE TEST PROCEDURE   *
*   IS STOPPED AS SOON AS THE MULTIPLE P-VALUE IS >= ALPHA FOR THE      *
*   FIRST TIME NO PROBLEM ARISES IF   ALPHA <= 1./FLOAT(KD).            *
*   THE STOPPING RULE FOR THE SEQUENTIAL MULTIPLE TEST PROCEDURE IS     *
*   THE FOLLOWING: STOP AS SOON AS FLOAT(IP) > FLOAT(NPERM)*ALPHA.      *
*************************************************************************

      DO 4 MM=1,NANZ-1
      IF(HILF(NANZ).EQ.HILF(MM)) THEN
        NANZ = NANZ - 1
        GOTO 5
      ENDIF
    4 CONTINUE

    5 CONTINUE

    6 CONTINUE

      IQUOT(MMM) = NANZ
      IF(ALPHA.EQ.1.000) GOTO 7

*************************************************************************
*   CHECK FOR STOPPING THE MULTIPLE TEST PROCEDURE.                     *
*************************************************************************

      IF(FLOAT(IQUOT(MMM)).GE.FLOAT(NPERM)*ALPHA) THEN
        DO 8 M=MMM,KD
        IQUOT(M)=-99
        RETURN
      END IF
      CONTINUE

*************************************************************************
*   THE KD-TH MULTIPLE P-VALUE IS IDENTICAL TO THE LARGEST UNI-         *
*   VARIATE P-VALUE.                                                    *
*************************************************************************

    8 IQUOT(KD) = IPGR(INDEX(KD)) + IPEQ(INDEX(KD))
      RETURN

      END
```

ANHANG B4

```
*************************************************************************
*                                                                       *
*                      PROGRAM SPLIT-PLOT MANOVA                        *
*                                                                       *
*            VERSION FOR THE MULTIVARIATE SPLIT-PLOT DESIGN             *
*                                                                       *
*                                                                       *
*                                                                       *
*     MC=1 : EXACT PERMUTATION TEST (PERFORMED IF TOTAL NUMBER OF       *
*            ADMISSIBLE PERMUTATIONS IS .LE. NPMAX (DEFAULT=20000)      *
*            PERMUTATIONS GENERATED IN SUBROUTINE KOMBIR                *
*                                                                       *
*     MC=0 : MONTE-CARLO SOLUTION USING NPERM <= NPMAX RANDOMLY         *
*            SELECTED ADMISSIBLE PERMUTATIONS DRAWN WITH REPLACEMENT    *
*            IN SUBROUTINE RANPER                                       *
*                                                                       *
*     NPMAX: MAXIMUM NUMBER OF ADMISSIBLE PERMUTATIONS FOR WHICH AN     *
*            EXACT PERMUTATION TEST IS STILL CARRIED OUT.               *
*            IF NPERM > NPMAX, A MONTE-CARLO SOLUTION USING NPMAX       *
*            RANDOM PERMUATIONS WITH REPLACEMENT WILL BE PERFORMED.     *
*                                                                       *
*     IR   : NUMBER OF GROUPS (PLOT-FACTOR LEVELS) (2<=IR<=10)          *
*                                                                       *
*     C    : NUMBER OF SPLIT-FACTOR LEVELS                              *
*                                                                       *
*     N(I) : SAMPLE SIZE FOR PLOT-FACTOR LEVEL I, I=1,IR                *
*                                                                       *
*     NSUM : N(1)+...+N(IR)  TOTAL SAMPLE SIZE  (NSUM <= 1000)          *
*                                                                       *
*     KD   : NUMBER OF DEPENDENT VARIABLES      (1 <= KD <= 10)         *
*                                                                       *
*     Y(L,C*K):L=1,NSUM; K=1,KD  RAW DATA READ FROM INPUT FILE FILEIN   *
*                                (LOGICAL UNIT 10)                      *
*                                                                       *
*     RAW-DATA HAVE TO BE GROUPED ACCORDING TO FACTOR LEVELS ALREADY.   *
*                                                                       *
*     PARAMETERS ARE INPUT INTERACTIVELY VIA TERMINAL                   *
*                                                                       *
*     PRINTED OUTPUT IS TO FILE FILEOUT (LOGICAL UNIT 20)               *
*                                                                       *
*                                                                       *
*                                                                       *
*     ALGORITHM DESIGNED AND PROGRAMMED BY                              *
*     ------------------------------------                              *
*     K.WILLMES & H.RATAJCZAK                                           *
*     DEPT. OF NEUROLOGY, RWTH AACHEN, PAUWELSSTR.                      *
*     D-5100 AACHEN, WEST GERMANY       15/06/85                        *
*                                                                       *
*                                                                       *
*     SUBROUTINES KOMBIR,RANPER PROGRAMMED BY                           *
*     ---------------------------------------                           *
*     N.PYHEL & G.HENN                                                  *
*     DEPT. OF MEDICAL STATISTICS, RWTH AACHEN                          *
*     1978 (UNPUBLISHED MANUSCRIPT)                                     *
*     MODIFIED BY H.RATAJCZAK, DEPT. OF NEUROLOGY, RWTH   06/85         *
*                                                                       *
*************************************************************************

      INTEGER N(10),IZ(1000),IPGR(10),IPEQ(10),IQUOT(10),C
      INTEGER CPUBEG,CPUEND,EQUAL
      REAL Y(1000,100),TY(1000,100),RISTAT(10)
      REAL RPSTAT(10),EPV(10),YR(1000,10),YC(10,10),YM(10),YPM(10,10)
      REAL YMR(10,10,10)
      REAL*8 MAXDIFF

C *********************************************************************
```

```
C          A R R A Y  -  D I M E N S I O N I N G                        *
C                                                                       *
C     N(IR),IZ(NSUM),IPGR(KD),IPEQ(KD),IQUOT(KD)                        *
C     Y(NSUM,KD),TY(NSUM,KD),RISTAT(KD),RPSTAT(KD),EPV(KD)              *
C     YR(NSUM,KD),YC(C,KD),YM(KD),YPM(IR,KD),YMR(IR,C,KD)               *
C ***********************************************************************

      CHARACTER*20 MCTYPE
      CHARACTER*60 NFMT,FILEIN,FILEOUT,TITLE

      INTEGER IN,OUT

      DATA YR,YC,YM,YPM,YMR/11210*0./

      CALL CPU_TIM(CPUBEG)

C ***********************************************************************
C     INPUT- AND OUTPUT-UNITS VIA TERMINAL                              *
C ***********************************************************************

      IN =5
      OUT=6

C ***********************************************************************
C     INTERACTIVE INPUT OF TITLE,NUMBER OF DEPENDENT VARIABLES,         *
C     DATA-FORMAT, AND INPUT/OUTPUT FILES                               *
C ***********************************************************************

20    WRITE(OUT,100)
      READ(IN,101) LENTITLE,TITLE
      IF(LENTITLE.GT.60) GOTO 20
21    WRITE(OUT,102)
      READ(IN,*) KD
      IF(KD.LT.2.OR.KD.GT.10) GOTO 21
22    WRITE(6,103)
      READ(IN,101) LENFILEIN,FILEIN
      OPEN(UNIT=10,NAME=FILEIN,READONLY,TYPE='OLD',ERR=23)
      GOTO 24
23    WRITE(OUT,*) '### ERROR OPENING INPUT-FILE ###'
      GOTO 22

24    WRITE(OUT,104)
      READ(5,101) LENNFMT,NFMT
      DO 25 I=1,LENNFMT
      IF(NFMT(I:I).EQ.'I'.OR.NFMT(I:I).EQ.'A') THEN
        WRITE(OUT,*) '### INPUT FORMAT MUST BE F-FORMAT ###'
        GOTO 24
      END IF
25    CONTINUE

26    WRITE(OUT,105)
      READ(IN,101) LENFILEOUT,FILEOUT
      OPEN(UNIT=20,NAME=FILEOUT,TYPE='NEW',ERR=27)
      GOTO 28
27    WRITE(OUT,*) '### ERROR OPENING OUTPUT-FILE ###'
      GOTO 26
28    WRITE(OUT,106)
      READ(IN,*) MC
      IF((MC-1)*MC.NE.0) GOTO 28
      IF(MC.EQ.1) GOTO 15
      WRITE(OUT,107)
      READ(IN,*) NPERM
      MCTYPE='MONTE-CARLO SOLUTION'

C ***********************************************************************
C     INPUT OF DESIGN PARAMETERS AND RAW-DATA                           *
```

```
C ***********************************************************************

   15 CALL SPIN(KD,NFMT,Y,TY,YR,YC,YM,YPM,YMR,N,IR,C,NSUM,EQUAL)

      WRITE(20,108) TITLE(1:LENTITLE)

C ***********************************************************************
C     OUTPUT OF DESIGN PARAMETERS AND RAW-DATA                          *
C ***********************************************************************

      CALL SPOUT(KD,NFMT,Y,TY,YR,YC,YM,YPM,YMR,N,IR,C,NSUM,FILEIN)

C ***********************************************************************
C     INITIALIZATION                                                    *
C                                                                       *
C     K1: SEED FOR THE RANDOM NUMBER GENERATOR RAN( ) IN SUBROUTINE     *
C     RANPER YIELDS DIFFERENT INITIAL VALUES FOR DIFFERENT RUNS OF THE  *
C     PROGRAM. IF IDENTICAL STARTS ARE REQUESTED SET K1= ANY INTEGER    *
C ***********************************************************************

      K1 = SECNDS(0.0)

C ***********************************************************************
C     EPS: MAXIMUM ABSOLUTE DIFFERENCE TAKEN TO BE ZERO                 *
C ***********************************************************************

      EPS = 1.E-4

C ***********************************************************************
C     NPMAX: MAXIMUM NUMBER OF PERMUTATIONS FOR EXACT OR MONTE-CARLO    *
C            SOLUTION                                                   *
C ***********************************************************************

      NPMAX = 20000

C ***********************************************************************
C     EXACT NUMBER OF PERMUTATION LESS EQUAL NPMAX ?                    *
C ***********************************************************************

      IF(MC.EQ.1) CALL ANZPERM(IR,N,NSUM,NPERM,NPMAX,MC,MCTYPE,EQUAL,
     -                                                          IKENN)

      IF(IKENN.EQ.1) THEN
        WRITE(20, 132) NPMAX,NPMAX
        WRITE(OUT,132) NPMAX,NPMAX
      END IF

C ***********************************************************************
C     COMPUTATION OF UNIVARIATE TEST-STATISTIC VALUES FOR EACH PERMU-   *
C     TATION STORED IN ARRAY VERTEILU(KD,NPMAX) AND COMPUTATION OF      *
C     UNIVARIATE P> AND P= VALUES FOR EACH OF THE KD DEPENDENT          *
C     VARIABLES                                                         *
C     SUBROUTINE SPLIT_PLOT ADEQUATE FOR A MULTIVARIATE SPLIT-PLOT      *
C     DESIGN                                                            *
C ***********************************************************************

      CALL SPLIT_PLOT(MC,NPERM,K1,EPS,KD,TY,IR,C,N,NSUM,IZ,IPGR,
     1IPEQ,RISTAT,RPSTAT,PVI,MPGR,MPEQ,MAXDIFF)

      WRITE(20,108) TITLE(1:LENTITLE)
      WRITE(20,137) MAXDIFF

C ***********************************************************************
C     OUTPUT OF RESULTS OF MULTIVARIATE TEST FOR DEPENDENT VAR. 1..KD   *
C ***********************************************************************
```

```
      WRITE(20,135) MCTYPE,NPERM
      WRITE(20,136) 1,KD,PVI,MPGR,MPEQ,FLOAT(MPGR+MPEQ)/FLOAT(NPERM)
      WRITE(20,142)

C *********************************************************************
C     OUTPUT OF RESULTS FOR UNIVARIATE TESTS FOR DEPENDENT VAR. 1..KD *
C *********************************************************************

      WRITE(20,133)

      DO 5 K=1,KD
      EPV(K) = FLOAT(IPGR(K)+IPEQ(K))/FLOAT(NPERM)
    5 WRITE(20,134) K,RISTAT(K),IPGR(K),IPEQ(K),EPV(K)

      WRITE(20,141)

C *********************************************************************
C     CALCULATION OF THE CPU-TIME REQUIRED                            *
C *********************************************************************

      CALL CPU_TIM(CPUEND)
      CPUTIME=FLOAT(CPUEND-CPUBEG)/100.
      WRITE(20,144) CPUTIME

C *********************************************************************
C     CLOSING THE INPUT- AND OUTPUT-UNIT                              *
C *********************************************************************

      CLOSE(10)
      CLOSE(20)

C --------------------------- FORMAT - STATEMENTS ----------------------
  100 FORMAT(' *** ENTER TITLE (MAX. 60 CHARACTERS)                  :',$)
  101 FORMAT(Q,A60)
  102 FORMAT(' *** ENTER NUMBER OF DEPENDENT VARIABLES (MAX.10)      :',$)
  103 FORMAT(' *** ENTER NAME OF INPUT FILE                          :',$)
  104 FORMAT(' *** ENTER INPUT FORMAT(MAX. 60 CHARACTERS,F-FORMAT)   :',$)
  105 FORMAT(' *** ENTER NAME OF OUTPUT FILE                         :',$)
  106 FORMAT(' *** ENTER 0 = MONTE-CARLO; 1 = EXACT SOLUTION         :',$)
  107 FORMAT(' *** ENTER NUMBER OF RANDOM PERMUTATIONS(<=NPMAX)      :',$)
  108 FORMAT(1H1,71(1H*)/
     -1X,1H*,16X,'      MANOVA PERMUTATION TEST      ',T72,1H*/
     -1X,1H*,11X,'FOR THE DEPENDENT VARIABLES OF A MANOVA-DESIGN',
     -T72,1H*/1X,1H*,69X,1H*/
     -1X,1H*,2X,'K.WILLMES & H.RATAJCZAK, DEPT. OF NEUROLOGY',
     -' RWTH AACHEN, GERMANY',T72,1H*/1X,1H*,69X,1H*/1X,1H*,14X,
     1'***      SPLIT-PLOT TEST PROBLEM      ***',T72,1H*/1X,1H*,T72,1H*
     1/1X,1H*,69X,1H*/1X,1H*,<(69-LENTITLE)/2>X,A<LENTITLE>,T72,1H*/
     11X,71(1H*)/)
  132 FORMAT(1H0,'*** EXACT SOLUTION NOT CARRIED OUT BECAUSE'/
     11X,'*** THE NUMBER OF PERMUTATIONS REQUIRED EXCEEDS THE'/
     21X,'*** MAXIMUM NUMBER OF NPMAX =',I6//
     31X,'*** THEREFORE A MONTE-CARLO SOLUTION USING'/
     41X,'*** NPERM =',I6,' PERMUTATIONS DRAWN RANDOMLY WITH'/
     51X,'*** REPLACEMENT IS CARRIED OUT'/1X,71(1H*)/)
  133 FORMAT(//
     114X,'RESULTS OF THE UNIVARIATE PERMUTATION TESTS :'/
     17X,'-------------------------------------------------------'/
     17X,'I  DEPENDENT  I    VALUE    I        I        I          I'/
     17X,'I  VARIABLE   I  TEST-STAT  I   P>   I   P=   I  P-VALUE  I'/
     17X,'I-------------+------------+--------+--------+----------+')
  134 FORMAT(7X,'I',4X,I2,7X,'I ',F11.4,' I ',I6,' I ',I6,' I  ',
     1       F7.4,'  I')
  135 FORMAT(8X,'*** ',A20,' USING ',I6,' PERMUTATIONS ***'////
     113X,'RESULTS OF THE MULTIVARIATE PERMUTATION TEST  :'/
     17X,'-------------------------------------------------------'/
```

```
     17X,'I  DEPENDENT  I PILLAI TRACE I      I        I          I'/
     17X,'I  VARIABLES  I  TEST-STAT   I   P>   I   P=   I  P-VALUE  I'/
     17X,'I-------------+--------------+--------+--------+----------+')
  136 FORMAT(7X,'I',2X,I2,' - ',I2,4X,'I ',F11.4,'  I ',I6,' I ',
     -       I6,' I  ',F7.4,'  I')
  137 FORMAT(//60X,' '/
     -' DIAGNOSTIC CHECK FOR ACCURACY OF GEN. INVERSE APLUS=',
     -'(SH+SE)  OF A=SH+SE:'//
     -'       ***  MAXDIFF ( A X APLUS X A  -  A ) : ',G12.5,' ***'///)
  141 FORMAT(7X,'I',13(1H-),'I',13(1H-),'I',2(8(1H-),'I'),11(1H-),'I')
  142 FORMAT(7X,'I',13(1H-),'I',14(1H-),'I',2(8(1H-),'I'),11(1H-),'I')
  144 FORMAT(12X,'CPUTIME :',F7.2,' SEC')
C ---------------------------------------------------------------------

      END

C **********************************************************************

      SUBROUTINE ANZPERM(IR,N,NSUM,NPERM,NPMAX,MC,MCTYPE,EQUAL,IKENN)

**************************************************************************
*     SUBROUTINE ANZPERM DETERMINES WHETHER THE NUMBER OF EXACT          *
*     PERMUTATIONS IS LESS THAN NPMAX. IF NOT, MCTYPE IS SET TO          *
*     MONTE-CARLO AND NPMAX PERMUTATIONS ARE CARRIED OUT                 *
**************************************************************************

      INTEGER N(10),EQUAL
      CHARACTER*20 MCTYPE

      MCTYPE='MONTE-CARLO SOLUTION'

      IKENN=0

C **********************************************************************
C     IDENTICAL PART FOR EQUAL OR UNEQUAL SAMPLE SIZES                   *
C **********************************************************************

      DO I=N(1)+1,NSUM
        ZAEHLER=ZAEHLER+ALOG10(FLOAT(I))
      END DO

      DO I=2,IR
        DO K=2,N(I)
          ZAEHLER=ZAEHLER-ALOG10(FLOAT(K))
        END DO
      END DO

C **********************************************************************
C     EQUAL SAMPLE SIZES                                                 *
C **********************************************************************
      IF(EQUAL.EQ.1.AND.IR.GE.2) THEN
        DO I=2,IR
          ZAEHLER=ZAEHLER-ALOG10(FLOAT(I))
        END DO
      END IF

C **********************************************************************
C     EXACT NUMBER OF PERMUTIONS GREATER THAN NPMAX                      *
C **********************************************************************

      IF(ZAEHLER.GT.ALOG10(FLOAT(NPMAX))) THEN
        IKENN=1
        NPERM=NPMAX
        MC=0
        MCTYPE='MONTE-CARLO SOLUTION'
      ELSE
```

```
        MCTYPE='   EXACT SOLUTION   '
        NPERM=INT(10.**ZAEHLER+0.5)
      END IF

      R E T U R N

      END

C ***************************************************************************

      SUBROUTINE SPLIT_PLOT(MC,NPERM,K1,EPS,KD,TY,IR,C,N,NSUM,IZ,
     1IPGR,IPEQ,RISTAT,RPSTAT,PVI,MPGR,MPEQ,MAXDIFF)

*****************************************************************************
*    SUBROUTINE SPLIT_PLOT CALLS THE SUBROUTINES NEEDED TO PERFORM THE      *
*    UNIVARIATE AND THE MULTIVARIATE PERMUTATION TESTS FOR THE              *
*    KD DEPENDENT VARIABLES                                                 *
*****************************************************************************

      INTEGER N(10),IPGR(10),IPEQ(10),IZ(1000),C,EIN,AUS,MEMO(1000)
      REAL RISTAT(10),RPSTAT(10)
      REAL TY(1000,100),SUM(10,10,10),TY_MEAN(10,10,10)
      REAL*8 SUMMAT(10,10),STOIN(10,10),MAXDIFF
      REAL SH(10,10),SE(10,10)
      LOGICAL START

C ***************************************************************************
C     INTEGER N(IR),IPGR(KD),IPEQ(KD),IZ(NSUM)                             *
C     REAL RISTAT(KD),VERTEILU(KD,NPERM),Y(NSUM,C*KD),RPSTAT(KD),          *
C    1TY(NSUM,C*KD)                                                        *
C ***************************************************************************

C ***************************************************************************
C     INITIALIZATION                                                       *
C ***************************************************************************

      MCC=(5-2*IR)*MC

      DO I=1,NSUM
        IZ(I)=I
      END DO

      NPERM1 = NPERM - 1
      START = .TRUE.
      DO 1 K=1,KD
        IPEQ(K) = 0
        IPGR(K) = 0
    1 RISTAT(K) = 0.

      MPGR=0
      MPEQ=0

C ***************************************************************************
C                  -                                                       *
C     COMPUTATION OF Ti.j                                                  *
C ***************************************************************************

      CALL GROUP_MEAN(TY,KD,IR,C,N,TY_MEAN,IZ)

      CALL HYPER(N,TY,IR,C,KD,SH,SE,TY_MEAN)

C ***************************************************************************
C                                                  -                       *
C     COMPUTATION OF THE PERMUTATION-INVARIANT PART (SH+SE)                *
C ***************************************************************************
```

```
      DO I=1,KD
        DO J=1,KD
          SUMMAT(I,J)=DBLE(SH(I,J)+SE(I,J))
        END DO
      END DO

      CALL G_INVERSE(SUMMAT,STOIN,KD,KD,MAXDIFF)

C *********************************************************************
C     COMPUTATION OF THE MANOVA TEST-STATISTIC FOR THE ORIGINAL      *
C     (UNPERMUTED) OBSERVATIONS IN PVI                               *
C *********************************************************************

      CALL PILSPUR(KD,PVI,STOIN,SH)

C *********************************************************************
C     COMPUTATION OF UNIVARIATE TEST STATISTIC VALUES FOR THE        *
C     ORIGINAL (UNPERMUTED) OBSERVATIONS  IN ARRAY RISTAT(KD)        *
C *********************************************************************

      CALL STAT_SP(KD,NSUM,IR,C,N,IZ,Y,TY,RISTAT,SUM,EIN,AUS)

      IF(MCC) 2,4,8

C *********************************************************************
C     AT EACH CALL KOMBIR GENERATES A SYSTEMATIC PERMUTATION OF THE  *
C     INTEGERS 1,NSUM IN ARRAY IZ(NSUM) AS LONG AS START=.FALSE.     *
C     AFTER THE LAST PERMUTATION IS GENERATED, START=.TRUE. IS       *
C     RETURNED                                                       *
C *********************************************************************

    2 CALL KOMBIR(START,IR,N,NSUM,IZ)

      IF(START) RETURN

      CALL GROUP_MEAN(TY,KD,IR,C,N,TY_MEAN,IZ)

C *********************************************************************
C     COMPUTATION OF MATRIX SH FOR PERMUTED DATA                     *
C *********************************************************************

      CALL HYPSH(N,IR,C,KD,SH,TY_MEAN)

C *********************************************************************
C     COMPUTATION OF MANOVA TEST-STATISTIC FOR PERMUTED DATA IN PVP  *
C *********************************************************************

      CALL PILSPUR(KD,PVP,STOIN,SH)

C *********************************************************************
C     TEST PVP > OR = PVI AND INCREASE COUNT IN MPGR OR MPEQ IF TRUE *
C *********************************************************************

      CALL MANOVA_PCOUNT(EPS,MPGR,MPEQ,PVI,PVP)

C *********************************************************************
C     COMPUTATION OF UNIVARIATE TEST STATISTIC VALUES FOR PERMUTED   *
C     DATA IN ARRAY RPSTAT(KD)                                       *
C *********************************************************************

      CALL STAT_SP(KD,NSUM,IR,C,N,IZ,Y,TY,RPSTAT,SUM,EIN,AUS)

C *********************************************************************
C     TEST RPSTAT( ) > OR = RISTAT( ) AND INCREASE COUNT IN IPGR( )  *
C      OR IPEQ( ) IF TRUE                                            *
C *********************************************************************
```

```
      CALL PCOUNT(KD,EPS,IPGR,IPEQ,RISTAT,RPSTAT)

      GOTO 2

C ***********************************************************************
C     PERFORM MONTE-CARLO SOLUTION                                      *
C ***********************************************************************

    4 DO 5 K=1,KD
    5 IPEQ(K) = 1
      MPEQ    = 1

C ***********************************************************************
C     GENERATE A RANDOM PERMUTATION OF 1,NSUM IN ARRAY IZ(NSUM) AS      *
C     LONG AS NPERM-1 CALLS ARE CARRIED OUT AND START=.TRUE.            *
C ***********************************************************************

    6 CALL RANPER(NPERM1,START,NSUM,IZ,K1)

      IF(START) RETURN

      CALL GROUP_MEAN(TY,KD,IR,C,N,TY_MEAN,IZ)

C ***********************************************************************
C     COMPUTATION OF MATRIX SH FOR PERMUTED DATA                        *
C ***********************************************************************

      CALL HYPSH(N,IR,C,KD,SH,TY_MEAN)

C ***********************************************************************
C     COMPUTATION OF MANOVA TEST-STATISTIC FOR PERMUTED DATA IN PVP     *
C ***********************************************************************

      CALL PILSPUR(KD,PVP,STOIN,SH)

C ***********************************************************************
C  TEST PVP > OR = PVI AND INCREASE COUNT IN MPGR OR MPEQ IF TRUE       *
C ***********************************************************************

      CALL MANOVA_PCOUNT(EPS,MPGR,MPEQ,PVI,PVP)

C ***********************************************************************
C     COMPUTATION OF UNIVARIATE TEST STATISTICS FOR THE RANDOMLY        *
C     PERMUTED DATA       ARRAY RPSTAT(KD)                              *
C ***********************************************************************

      CALL STAT_SP(KD,NSUM,IR,C,N,IZ,Y,TY,RPSTAT,SUM,EIN,AUS)

C ***********************************************************************
C     TEST RPSTAT( ) > OR = RISTAT( ) AND INCREASE COUNT IN IPGR( )     *
C     OR IPEQ( ) IF TRUE                                                *
C ***********************************************************************

      CALL PCOUNT(KD,EPS,IPGR,IPEQ,RISTAT,RPSTAT)

      GOTO 6

8     CALL KOMBI2(N(1),NSUM,START,IZ,MEMO,EIN,AUS)

      IF(START) RETURN

      CALL GROUP_MEAN(TY,KD,IR,C,N,TY_MEAN,IZ)

C ***********************************************************************
C     COMPUTATION OF MATRIX SH FOR PERMUTED DATA                        *
```

```
C ***********************************************************************

      CALL HYPSH(N,IR,C,KD,SH,TY_MEAN)

C ***********************************************************************
C     COMPUTATION OF MANOVA TEST-STATISTIC FOR PERMUTED DATA IN PVP  *
C ***********************************************************************

      CALL PILSPUR(KD,PVP,STOIN,SH)

C ***********************************************************************
C     TEST PVP > OR = PVI AND INCREASE COUNT IN MPGR OR MPEQ IF TRUE *
C ***********************************************************************

      CALL MANOVA_PCOUNT(EPS,MPGR,MPEQ,PVI,PVP)

C ***********************************************************************
C     COMPUTATION OF UNIVARIATE TEST STATISTICS FOR THE RANDOMLY      *
C     PERMUTED DATA       ARRAY RPSTAT(KD)                            *
C ***********************************************************************

      CALL EXACT2(KD,NSUM,IR,C,N,IZ,Y,TY,RPSTAT,SUM,EIN,AUS)

C ***********************************************************************
C     TEST RPSTAT( ) > OR = RISTAT( ) AND INCREASE COUNT IN IPGR( )  *
C     OR IPEQ( ) IF TRUE                                              *
C ***********************************************************************

      CALL PCOUNT(KD,EPS,IPGR,IPEQ,RISTAT,RPSTAT)

      GOTO 8

      END

C ***********************************************************************

      SUBROUTINE STAT_SP(KD,NSUM,IR,C,N,IZ,Y,TY,RSTAT,SUM,EIN,AUS)

***************************************************************************
*  IN SUBROUTINE STAT_SP THE KD UNIVARIATE TEST STATISTIC VALUES FOR A *
*  MULTIVARIATE SPLIT-PLOT DESIGN ARE COMPUTED BOTH FOR THE ORIGINAL   *
*  AND THE PERMUTED DATA                                               *
***************************************************************************

      REAL Y(1000,100),TY(1000,100),SUM(10,10,10),RSTAT(10)
      INTEGER N(10),IZ(1000),C,EIN,AUS

C ***********************************************************************
C     REAL Y(NSUM,C*KD),TY(NSUM,C*KD),SUM(IR,KD),RSTAT(KD)            *
C     INTEGER N(IR),IZ(NSUM)                                          *
C ***********************************************************************

C ***********************************************************************
C     COMPUTATION OF TEST-STATISTIC FOR EXACT AND MONTE_CARLO SOLUTION  *
C ***********************************************************************

      DO 10 K=1,KD
        RSTAT(K)=0.
        DO 10 J=1,C
        NZAEHL=0
        DO 10 I=1,IR
        SUM(I,J,K)=0.
        DO 20 L=1,N(I)
          NZAEHL=NZAEHL+1
20        SUM(I,J,K)=SUM(I,J,K)+TY(IZ(NZAEHL),K+(J-1)*KD)
        SUM(I,J,K)=SUM(I,J,K)/FLOAT(N(I))
```

```
10    RSTAT(K)=RSTAT(K)+FLOAT(N(I))*SUM(I,J,K)*SUM(I,J,K)

      R E T U R N

C *************************************************************************
C     COMPUTATION OF TEST-STATISTIC FOR EXACT SOLUTION WITH 2 PLOT-      *
C     FACTOR LEVELS                                                      *
C *************************************************************************

      ENTRY EXACT2(KD,NSUM,IR,C,N,IZ,Y,TY,RSTAT,SUM,EIN,AUS)

      DO 30 K=1,KD
      RSTAT(K)=0.
      DO 30 J=1,C
      SUM(1,J,K)=SUM(1,J,K)+(TY(EIN,K+(J-1)*KD)-TY(AUS,K+(J-1)*KD))/
     -                                                    FLOAT(N(1))
      SUM(2,J,K)=SUM(2,J,K)+(TY(AUS,K+(J-1)*KD)-TY(EIN,K+(J-1)*KD))/
     -                                                    FLOAT(N(2))
      DO 30 I=1,2
30    RSTAT(K)=RSTAT(K)+FLOAT(N(I))*SUM(I,J,K)*SUM(I,J,K)

      R E T U R N

      END

C *************************************************************************

      SUBROUTINE PCOUNT(KD,EPS,IPGR,IPEQ,RISTAT,RPSTAT)

*************************************************************************
*   SUBROUTINE PCOUNT COMPARES THE ORIGINAL UNIVARIATE TEST STATISTIC   *
*   VALUES WITH THE VALUES OBTAINED FROM EACH ADMISSIBLE PERMUTATION    *
*   THE COUNTS FOR P> OR P= ARE INCREASED IF RISTAT( ) > OR = RPSTAT(   *
*************************************************************************

      INTEGER IPGR(10),IPEQ(10)
      REAL RISTAT(10),RPSTAT(10)

C *************************************************************************
C     INTEGER IPGR(KD),IPEQ(KD)                                          *
C     REAL RISTAT(KD),RPSTAT(KD)                                         *
C *************************************************************************

C *************************************************************************
C     EPS IS THE MAXIMUM DIFFERENCE BETWEEN TEST STATISTIC VALUES        *
C     TAKEN TO BE ZERO                                                   *
C *************************************************************************

      DO 1 K=1,KD
      IF(ABS(RISTAT(K) - RPSTAT(K)).LT.EPS) IPEQ(K) = IPEQ(K) + 1
      IF((RPSTAT(K) - RISTAT(K)).GT.EPS) IPGR(K) = IPGR(K) + 1
    1 CONTINUE

      RETURN

      END

C *************************************************************************

      SUBROUTINE MANOVA_PCOUNT(EPS,MPGR,MPEQ,PVI,PVP)

*************************************************************************
*   SUBROUTINE PCOUNT COMPARES THE ORIGINAL MANOVA TEST STATISTIC       *
*   VALUE WITH THE VALUES OBTAINED FROM EACH ADMISSIBLE PERMUTATION.    *
*   THE COUNTS FOR P> OR P= ARE INCREASED IF PVI > OR = PVP             *
*************************************************************************
```

```
C ************************************************************************
C     EPS IS THE MAXIMUM DIFFERENCE BETWEEN TEST STATISTIC VALUES      *
C     TAKEN TO BE ZERO                                                 *
C ************************************************************************

      IF(ABS(PVI-PVP).LT.EPS) MPEQ = MPEQ + 1
      IF((PVP - PVI).GT.EPS)  MPGR = MPGR + 1

      RETURN

      END

C ************************************************************************

       SUBROUTINE KOMBI2 (K,N,START,Z,MEMO,IN,OUT)

C ************************************************************************
C      SUBROUTINE KOMBI2 GENERATES ALL ADMISSIBLE PERMUTATIONS FOR THE *
C      EXACT SOLUTION OF AN INDEPENDENT TWO-SAMPLES PROBLEM            *
C      THE PERMUTED INDECES ARE RETURNED IN ARRAY Z(N)                 *
C ************************************************************************

       INTEGER Z(1000),MEMO(1000),IN,OUT
       LOGICAL START

       IF (.NOT.START) GOTO 30
       WRITE(1,*)'EXAKTE LOESUNG'
       START=.FALSE.
       OUT=1
       IN=1

       DO 10 I=1,K
 10    MEMO(I)=0

C ************************************************************************
C      AT FIRST CALL RETURN ORIGINAL SEQUENCE OF DATA I.E. IDENTICAL   *
C      PERMUTATION                                                     *
C ************************************************************************

       DO 20 I=1,N
 20    Z(I)=I
       RETURN

 30    DO                          40 I=1,K
       L=K-I+1
       IF (MEMO(L).LT.N-K) GOTO 50
 40    CONTINUE
       START=.TRUE.
       RETURN
 50    M=MEMO(L)+1
       DO 60 I=L,K
 60    MEMO(I)=M
       OUT=Z(L)
       Z(L)=Z(M+K)
       IN=Z(L)
       Z(M+K)=OUT

       RETURN

       END

C ************************************************************************

      SUBROUTINE KOMBIR (START,R,N,NSUM,Z)
```

```
*************************************************************************
*    SUBROUTINE KOMBIR GENERATES ALL SYSTEMATIC PERMUTATIONS (ONE AT    *
*    EACH CALL) OF NSUM ELEMENTS IN R GROUPS OF N(I),I=1,R ELEMENTS     *
*    EACH.                                                              *
*                                                                       *
*    ALGORITHM PROGRAMMED BY N. PYHEL & G. HENN                         *
*************************************************************************

      INTEGER MEMO(10,1000),NEMO(1000),Z(1000),N(10),NPS(10),R,RM1,C,OUT
      LOGICAL START

C *************************************************************************
C     INTEGER MEMO(IR-1,NSUM),NEMO(NSUM),Z(NSUM),N(IR),NPS(IR)           *
C *************************************************************************

      IF (.NOT.START) GOTO  4

      DO 100 I=1,NSUM
100   NEMO(I)=0
      LEQ = 1
      DO 1 I = 1,R
      NPS(I) = 0
      DO 1 J = 1,NSUM
      MEMO(I,J) = 0
    1 CONTINUE
      START=.FALSE.
      RM1=R-1
      DO  2 I=1,R
      IF (N(1).NE.N(I)) LEQ=0
      DO  2 J=1,I
 2    NPS(I)=NPS(I)+N(J)
      DO 3 I=1,NSUM
      Z(I)=I
      DO 3 J=1,RM1
 3    MEMO(J,I)=I
      RETURN
 4    M=R
 5    M=M-1
      IF (M.GT.0) GOTO 6
      START=.TRUE.
      RETURN
 6    N1=N(M)-LEQ
      N2=NSUM-NPS(M)
      N3=N2+N(M)
      N4=NSUM-N3
      N5=NPS(M)
      DO 7 I=1,N1
      L=N5-I+1
      IF (NEMO(L).LT.N2) GOTO 9
 7    CONTINUE
      DO 8 I=1,N3
      NEMO(N4+I)=0
 8    Z(N4+I)=MEMO(M,N4+I)
      GOTO 5
 9    J=NEMO(L)+1
      DO 10 I=L,N5
10    NEMO(I)=J
      OUT=Z(L)
      Z(L)=Z(J+N5)
      Z(J+N5)=OUT
      IF (M.EQ.R-1) RETURN
      MP1=M+1
      DO 11 I=MP1,RM1
      DO 11 K=L,NSUM
 11   MEMO (I,K)=Z(K)
```

```
      RETURN

      END

C *************************************************************************

      SUBROUTINE RANPER (NP,START,N,Z,K1)

*************************************************************************
C     IN SUNROUTINE RANPER A RANDOM PERMUTATION IS GENERATED AT EACH    *
C     CALL USING THE UNIFORM RANDOM NUMBER GENERATOR RAN( ) WITH        *
C     SEEDING PARAMETER K1. IN TOTAL NP PERMUTATIONS WITH REPLACEMENT   *
C     ARE GENERATED. IF K1 IS A CONSTANT THE SAME SEQUENCE OF RANDOM    *
C     NUMBERS IS GENERATED AT EACH NEW RUN OF THE PROGRAM. IF K1 IS E.G.*
C     TIED TO THE REAL-TIME AT PROGRAM START, DIFFERENT SEQUENCES ARE   *
C     VERY LIKELY AT DIFFERENT RUNS.                                    *
*************************************************************************

      INTEGER Z(1000)
      LOGICAL START

      IF(.NOT.START) GOTO 1

      START=.FALSE.
      J=0

    1 DO 2 I=1,N
    2 Z(I)=I
      J=J+1
      IF (J.GT.NP) START=.TRUE.
      DO 3 I=1,N
      L=I+RAN(K1)*(N+1-I)
      K=Z(L)
      Z(L)=Z(I)
    3 Z(I)=K

      RETURN

      END

C *************************************************************************

      SUBROUTINE SPIN(KD,NFMT,Y,TY,YR,YC,YM,YPM,YMR,N,IR,C,NSUM,EQUAL)

*************************************************************************
C  IN SUBROUTINE SPIN THE DESIGN PARAMETERS ARE INPUT INTERACTIVELY     *
C  AND THE RAW DATA READ IN FROM LOGICAL UNIT 10                        *
*************************************************************************

      INTEGER N(10),EQUAL,C
      REAL Y(1000,100),TY(1000,100),YR(1000,10),YC(10,10),YM(10)
      REAL YPM(10,10),YMR(10,10,10)
      CHARACTER*60 NFMT

C *************************************************************************
C     INTEGER N(IR)                                                      *
C     REAL Y(NSUM,C*KD),TY(NSUM,C*KD),YR(NSUM,KD),YC(C,KD),YM(KD)        *
C     REAL YPM(R,KD),YMR(R,C,KD)                                         *
C *************************************************************************

      TYPE 100
      ACCEPT *,IR
      TYPE 108
      ACCEPT *,C
5     TYPE 105
      ACCEPT *,EQUAL
```

```
      IF((EQUAL-1)*(EQUAL-2).NE.0) GOTO 5
      IF(EQUAL.EQ.2) THEN
        TYPE 101
        ACCEPT *,(N(I),I=1,IR)
      ELSE
        TYPE 106
        ACCEPT *,N(1)
        DO I=2,IR
          N(I)=N(1)
        END DO
      END IF

      NSUM = 0
      DO 1 I=1,IR
      NSUM = NSUM + N(I)
      IF(NSUM.GT.1000) THEN
        TYPE 107
        STOP
      END IF
1     CONTINUE

C *********************************************************************
C     READ RAW-DATA FROM LOGICAL UNIT 10 USING FORMAT NFMT          *
C *********************************************************************

      M=C*KD
      NZAEHL=0
      DO NR=1,IR
        DO L=1,N(NR)
          NZAEHL=NZAEHL+1
          READ(10,NFMT,ERR=3,END=4) (Y(NZAEHL,K),K=1,M)

          DO J=1,C
            DO K=1,KD
              YC (J,K)      = YC (J,K)      +Y(NZAEHL,K+(J-1)*KD)
              YR (NZAEHL,K)= YR (NZAEHL,K)+Y(NZAEHL,K+(J-1)*KD)
              YM (K)        = YM (K)        +Y(NZAEHL,K+(J-1)*KD)
              YPM(NR,K)     = YPM(NR,K)     +Y(NZAEHL,K+(J-1)*KD)
              YMR(NR,J,K)   = YMR(NR,J,K)   +Y(NZAEHL,K+(J-1)*KD)
            END DO
          END DO
        END DO
      END DO

      DO K=1,KD
        YM(K)=YM(K)/FLOAT(C*NSUM)

          DO NZAEHL=1,NSUM
            YR(NZAEHL,K)=YR(NZAEHL,K)/FLOAT(C)
          END DO

          DO NR=1,IR
            YPM(NR,K)=YPM(NR,K)/FLOAT(C*N(NR))
          END DO

          DO J=1,C
            YC(J,K)=YC(J,K)/FLOAT(NSUM)

            DO NR=1,IR
              YMR(NR,J,K)=YMR(NR,J,K)/FLOAT(N(NR))
            END DO
          END DO
        END DO

C *********************************************************************
C     COMPUTATION OF TRANSFORMED VALUES TY                           *
```

```
C ***********************************************************************

      DO L=1,NSUM
        DO J=1,C
          DO K=1,KD
            TY(L,K+(J-1)*KD)=Y(L,K+(J-1)*KD)-YR(L,K)-YC(J,K)+YM(K)
          END DO
        END DO
      END DO

      TYPE 102
      RETURN

    3 TYPE 103
      STOP

    4 TYPE 104
      STOP

C ---------------------- FORMAT - STATEMENTS ----------------------------
  100 FORMAT(' *** MANOVA SPLIT - PLOT DESIGN ***'//
     1' *** ENTER PARAMETERS ***'//
     2' *** 1. NUMBER OF  PLOT LEVELS (MAX.10)              :',$)
  101 FORMAT(' *** 3. SAMPLE SIZES N(1),..,N(R) (TOTAL SIZE<=1000) :',$)
  102 FORMAT(//' *** RAW DATA READ IN ***'/)
  103 FORMAT(//' *** ERROR DETECTED WHEN READING IN RAW DATA ***'/)
  104 FORMAT(//' *** INSUFFICIENT RAW DATA ***'/)
  105 FORMAT(' *** 3. EQUAL (1) OR UNEQUAL (2) SAMPLE SIZES         :',$)
  106 FORMAT(' *** 4. EQUAL SAMPLE SIZE FOR ALL <IR> SAMPLES        :',$)
  107 FORMAT(//' ### SUM OF SAMPLE SIZES > 1000 !! ###')
  108 FORMAT(' *** 2. NUMBER OF SPLIT LEVELS (MAX.10)               :'
     -,$)

C ---------------------------------------------------------------------

      END

C ***********************************************************************

      SUBROUTINE SPOUT(KD,NFMT,Y,TY,YR,YC,YM,YPM,YMR,N,IR,C,NSUM,FILEIN)

C ***********************************************************************
C     IN SUBROUTINE SPOUT THE RAW DATA OF A SPLIT-PLOT DESIGN ARE      *
C     OUTPUT TO LOGICAL UNIT 20                                        *
C ***********************************************************************

      REAL Y(1000,100),TY(1000,100),YR(1000,10),YC(10,10),YM(10)
      REAL YPM(10,10),YMR(10,10,10)
      INTEGER N(10),C

C ***********************************************************************
C     REAL Y(NSUM,C*KD)                                                *
C     INTEGER N(IR),C                                                  *
C ***********************************************************************

      CHARACTER*60 FILEIN,NFMT

      WRITE(20,100) FILEIN,NFMT

      DO K=1,KD
        IF(K.GT.1) WRITE(20,106)
        WRITE(20,101) K,(J,J=1,C)
        NZAEHL=0
        DO L=1,IR
          DO I=1,N(L)
            NZAEHL=NZAEHL+1
```

```
            IF(I.NE.(N(L)+1)/2) THEN
             WRITE(20,102) I,(Y(NZAEHL,K+(J-1)*KD),J=1,C),YR(NZAEHL,K)
            ELSE
             WRITE(20,103) L,I,(Y(NZAEHL,K+(J-1)*KD),J=1,C),YR(NZAEHL,K)
            END IF
          END DO
          WRITE(20,104) (YMR(L,J,K),J=1,C),YPM(L,K)
        END DO
        WRITE(20,105) (YC(J,K),J=1,C),YM(K)
      END DO

C *************************************************************************
C     OUTPUT OF TRANSFORMED VALUES TY                                      *
C *************************************************************************

      DO K=1,KD
        WRITE(20,206)
        WRITE(20,201) K,(J,J=1,C)
        NZAEHL=0
        DO L=1,IR
          DO I=1,N(L)
            NZAEHL=NZAEHL+1
            IF(I.NE.(N(L)+1)/2) THEN
              WRITE(20,202) I,(TY(NZAEHL,K+(J-1)*KD),J=1,C)
            ELSE
              WRITE(20,203) L,I,(TY(NZAEHL,K+(J-1)*KD),J=1,C)
            END IF
          END DO
          WRITE(20,204)
        END DO
      END DO

      RETURN

C ---------------------------- FORMAT - STATEMENTS ---------------------

  100 FORMAT(25X,21(1H*)/25X,1H*,' R A W  -  D A T A ',1H*/25X,
     221(1H*)//1X,'INPUT FILE  : ',A60/1X,'DATA-FORMAT : ',A60/)
  101 FORMAT(1X,16(1H-),'---',<C>(10(1H-)),'----------------'/
     -1X,'I  DEPENDENT    ','I',<(C-2)*5+3>X,'   SPLIT  FACTOR   ',
     -<(C-2)*5+2>X,'I','    BLOCK   I'/
     -1X,'I  VARIABLE  ',I2,1X,'I',2X,<C>(4X,I2,4X),2X,'I    MEAN    I'/
     -1X,16(1H-),'I--',<C>(10(1H-)),'--I------------I')
  102 FORMAT(1X,'I',10X,'I',I3,' I',2X,<C>F10.3,'  I',F10.3,'  I')
  103 FORMAT(1X,'I  PLOT  ',I1,' I',I3,' I',2X,<C>F10.3,'  I',F10.3,
     -       '  I')
  104 FORMAT(1X,'I----------I----I--',<C>(10(1H-)),'--I------------I'/
     -       1X,'I  MEAN    I    I',2X,<C>F10.3,'  I',F10.3,'  I',
     -       '  PLOT MEAN'/
     -       1X,'I----------I----I--',<C>(10(1H-)),'--I------------I')
  105 FORMAT(1X,'I  SPLIT   I    I  ',<C*10>X,'  I            I',
     -       '  OVERALL'/
     -       1X,'I  MEAN    I    I  ',<C>F10.3,'  I',F10.3,'  I  MEAN'/
     -       1X,'I------------------',<C>(10(1H-)),'--I------------I')
  106 FORMAT(1H1)
  201 FORMAT(1X,16(1H-),'---',<C>(10(1H-)),'---'/
     -1X,'I  DEPENDENT    ','I',<(C-2)*5+3>X,'   SPLIT  FACTOR   ',
     -<(C-2)*5+2>X,'I'/
     -1X,'I  VARIABLE  ',I2,1X,'I',2X,<C>(4X,I2,4X),2X,'I'/
     -1X,16(1H-),'I--',<C>(10(1H-)),'--I')
  202 FORMAT(1X,'I',10X,'I',I3,' I',2X,<C>F10.3,'  I')
  203 FORMAT(1X,'I  PLOT  ',I1,' I',I3,' I',2X,<C>F10.3,'  I')
  204 FORMAT(1X,'I----------I----I--',<C>(10(1H-)),'--I')
  206 FORMAT(1H1,'TRANSFORMED VALUES TY'//)
C ---------------------------------------------------------------------
```

```
      END
C ***************************************************************************

      SUBROUTINE GROUP_MEAN(TY,KD,IR,C,N,TY_MEAN,IZ)

C ***************************************************************************
C                    -                                                      *
C *** COMPUTATION OF Ti.j                                                   *
C *** TY_MEAN (I,J,K)  I = 1,IR,  J=1,C,   K = 1,KD                         *
C                                                                           *
C     I N P U T:                                                            *
C     KD          -  DIMENSION OF VECTORS                                   *
C     IR          -  NUMBER OF PLOT-FACTOR LEVELS (GROUPS)                  *
C     C           -  NUMBER OF SPLIT-FACTOR LEVELS                          *
C     N(I)        -  SAMPLE SIZE FOR PLOT-LEVEL I, I=1,..,IR                *
C     TY(I,K)     -  OBSEVATION VECTORS  I = 1,NSUM  K = 1,KD               *
C                                                                           *
C    O U T P U T:                                                           *
C    TY_MEAN(I,K) -  VECTOR OF GROUP-MEANS PER SPLIT-FACTOR LEVEL           *
C                                                                           *
C ***************************************************************************

      REAL TY(1000,100),TY_MEAN(10,10,10)
      INTEGER N(10),IZ(1000),C

      DO K=1,KD
        DO J=1,C
          NZAEHL=0
          DO I=1,IR
            TY_MEAN(I,J,K)=0.
            DO L=1,N(I)
              NZAEHL=NZAEHL+1
              TY_MEAN(I,J,K)=TY_MEAN(I,J,K)+TY(IZ(NZAEHL),K+(J-1)*KD)
            END DO
            TY_MEAN(I,J,K)=TY_MEAN(I,J,K)/FLOAT(N(I))
          END DO
        END DO
      END DO

      R E T U R N

      END

C ***************************************************************************

      SUBROUTINE HYPER (N,TY,IR,C,KD,SH,SE,TY_MEAN)

C***************************************************************************
C                                                                           *
C     COMPUTATION OF THE KD X KD - MATRICES SH AND SE                       *
C                                                                           *
C     I N P U T:                                                            *
C     N(I)           -  SAMPLE SIZE FOR GROUP I=1,..,NR                     *
C     TY(I,K)        -  TRANSFORMED DATA                                    *
C     IR             -  NUMBER OF PLOT-FACTOR LEVELS                        *
C     C              -  NUMBER OF SPLIT-FACTOR LEVELS                       *
C     KD             -  NUMBER OF DEPENDENT VARIABLES                       *
C     TY_MEAN(I,J,K)-  GROUP MEANS PER SPLIT-FACTOR LEVEL                   *
C                                                                           *
C     O U T P U T:                                                          *
C     SH(K,L); SE(K,L) - MATRICES SH AND SE: K = 1,KD; L = 1,KD             *
C                                                                           *
C***************************************************************************

      REAL TY(1000,100),SE(10,10),SH(10,10),TY_MEAN(10,10,10)
```

```
      INTEGER N(10),C

      DO K=1,KD
        DO L=1,KD
          SH(K,L)=0.
          SE(K,L)=0.
          DO J=1,C
            NZAEHL=0
            DO I=1,IR
              SH(K,L)=SH(K,L)+N(I)*TY_MEAN(I,J,K)*TY_MEAN(I,J,L)
              DO M=1,N(I)
                NZAEHL=NZAEHL+1
                SE(K,L)=SE(K,L)+TY(NZAEHL,K+(J-1)*KD)*TY(NZAEHL,L+(J-1)*KD)
              END DO
            END DO
          END DO
          SE(K,L)=SE(K,L)-SH(K,L)
        END DO
      END DO

      R E T U R N

      END

C *************************************************************************

      SUBROUTINE HYPSH (N,IR,C,KD,SH,TY_MEAN)

C *************************************************************************
C                                                                         *
C      COMPUTATION OF THE KD X KD MATRIX  SH                              *
C                                                                         *
C      I N P U T:                                                         *
C      N(I)           -  SAMPLE SIZE FOR GROUP I=1,..,IR                  *
C      TY(I,K)        -  TRANSFORMED DATA                                 *
C      IR             -  NUMBER OF PLOT-LEVELS                            *
C      C              -  NUMBER OF SPLIT-LEVELS                           *
C      KD             -  NUMBER OF DEPENDENT VARIABLES                    *
C      TY_MEAN(I,J,K)-  GROUP MEANS PER SPLIT-LEVEL                       *
C                                                                         *
C      O U T P U T:                                                       *
C      SH(K,L);  - MATRIX SH : K = 1,KD; L = 1,KD                         *
C                                                                         *
C*************************************************************************

      REAL SH(10,10),TY_MEAN(10,10,10)
      INTEGER N(10),C

      DO K=1,KD
        DO L=1,KD
        SH(K,L)=0.
          DO J=1,C
            DO I=1,IR
              SH(K,L)=SH(K,L)+N(I)*TY_MEAN(I,J,K)*TY_MEAN(I,J,L)
            END DO
          END DO
        END DO
      END DO

      R E T U R N

      END

C *************************************************************************

        SUBROUTINE G_INVERSE ( A , APLUS , NZEIL , NSPALT , MAXDIFF )
```

```
*************************************************************************
*                                                                       *
*     SUBROUTINE FOR COMPUTATION OF THE MOORE-PENROSE GENERALIZED INVERSE *
*       USING THE ALGORITHM OF BOUILLON,T.L. & ODELL,P.L.(1971).        *
*       GENERALIZED INVERSE MATRICES. NEW YORK: WILEY.                  *
*                                                                       *
*     ALGORITHM PROGRAMMED BY HERBERT RATAJCZAK , 6/85                  *
*                                                                       *
*************************************************************************


      REAL*8 A(10,10),APLUS(10,10),EPS,SUMME,D(10),C(10),B(10)
      REAL*8 Z1(10,10),Z2(10,10),MAXDIFF

C *************************************************************************
C     INITIALIZE                                                          *
C                                                                         *
C     EPS: ACCURACY LEVEL                                                 *
C *************************************************************************

      DATA EPS /1.0E-8/
      DATA B,C,D /30*0./
      DO I=1,NZEIL
        DO K=1,NSPALT
          APLUS(K,I)=0.
        END DO
      END DO

C *************************************************************************
C     INITIALIZATION OF THE ALGORITHM                                     *
C                                                                         *
C     COMPUTATION OF A1+                                                  *
C     ------------------                                                  *
C                                                                         *

      SUMME=0.
      DO I=1,NZEIL
      SUMME=SUMME+A(I,1)
      END DO

C *************************************************************************
C     A1+ = 0 ?                                                           *
C *************************************************************************

      IF(SUMME.GT.EPS) THEN

      SUMME=0.
      DO I=1,NZEIL
      SUMME=SUMME+A(I,1)*A(I,1)
      END DO

      IF(SUMME.NE.0) SUMME=1./SUMME

      DO I=1,NZEIL
      APLUS(1,I)=A(I,1)*SUMME
      END DO

      END IF

      INDEX=1

C *************************************************************************
C     ITERATION                                                           *
C     =========                                                           *
C *************************************************************************
```

```
      DO WHILE ( INDEX . LE . NSPALT )

      INDEX=INDEX+1

C *********************************************************************
C       COMPUTATION OF THE D(I)                                       *
C       -----------------------                                       *
C *********************************************************************

      DO I=1,INDEX-1
      D(I)=0.
      DO K=1,NZEIL
      D(I)=D(I)+APLUS(I,K)*A(K,INDEX)
      END DO
      END DO

C *********************************************************************
C       COMPUTATION OF THE C(I)                                       *
C       -----------------------                                       *
C *********************************************************************

      SUMME=0.
      DO K=1,NZEIL
      C(K)=A(K,INDEX)
      DO I=1,INDEX-1
      C(K)=C(K)-A(K,I)*D(I)
      END DO
      SUMME=SUMME+C(K)
      END DO

C *********************************************************************
C       COMPUTATION OF THE B(I)                                       *
C       -----------------------                                       *
C *********************************************************************

      IF(DABS(SUMME).GT.EPS) THEN

      SUMME=0.
      DO I=1,NZEIL
      SUMME=SUMME+C(I)*C(I)
      END DO
      IF(SUMME.NE.0) SUMME=1./SUMME

      DO I=1,NZEIL
      B(I)=C(I)*SUMME
      END DO

      ELSE

      SUMME=1.
      DO I=1,INDEX-1
      SUMME=SUMME+D(I)*D(I)
      END DO
      IF(SUMME.NE.0) SUMME=1./SUMME

      DO K=1,NZEIL
      B(K)=0.
      DO I=1,INDEX-1
      B(K)=B(K)+SUMME*D(I)*APLUS(I,K)
      END DO
      END DO

      END IF

C *********************************************************************
C       COMPUTATION OF APLUS                                          *
```

```
C       --------------------                                            *
C ***************************************************************************

          DO I=1,INDEX
          DO K=1,NZEIL
          IF(I.NE.INDEX) THEN
          APLUS(I,K)=APLUS(I,K)-D(I)*B(K)

          ELSE

          APLUS(I,K)=B(K)
          END IF
          END DO
          END DO

          END DO

C ***************************************************************************
C       COMPUTATION OF THE MAX. DIFFERENCE A X APLUS X A  -  A             *
C       ----------------------------------------------------               *
C ***************************************************************************

          CALL PRODUKT(A,APLUS,Z1,NZEIL,NSPALT,NSPALT,NZEIL)
          CALL PRODUKT(Z1,A,Z2,NZEIL,NZEIL,NZEIL,NSPALT)

          MINDIFF=0.0

          DO I=1,NZEIL
          DO K=1,NSPALT
          IF(DABS(A(I,K)-Z2(I,K)).GT.MINDIFF) MINDIFF=DABS(A(I,K)-Z2(I,K))
          END DO
          END DO

          R E T U R N

          E N D

C ***************************************************************************

          SUBROUTINE PRODUKT (X,Y,Z,N1,M1,N2,M2)

***************************************************************************
*       THE SUBROUTINE COMPUTES THE PRODUCT  X*Y=Z                        *
***************************************************************************

          IMPLICIT REAL*8 (A-H,O-Z)
          REAL*8 X(10,10),Y(10,10),Z(10,10)

          IF(M1.NE.N2) THEN
          WRITE(6,*) '#### MATRICES CANNOT BE MULTIPLIED ####'
          STOP
          END IF

C ***************************************************************************
C       COMPUTATION OF THE MATRIX PRODUCT                                  *
C ***************************************************************************

          DO I=1,N1
          DO K=1,M2
          Z(I,K)=0.D00
          DO L=1,M1
          Z(I,K)=Z(I,K)+X(I,L)*Y(L,K)
          END DO
          END DO
```

```
      END DO

      R E T U R N

      END

C *********************************************************************

      SUBROUTINE PILSPUR(KD,PV,STOIN,SH)

C *********************************************************************
C                                                                      *
C     SUBROUTINE FOR THE COMPUTATION OF A MANOVA TEST STATISTIC        *
C               USING PILLAI'S TRACE CRITERION                         *
C                                                                      *
C *********************************************************************

      REAL*8 STOIN(10,10),SPUMA(10,10)
      REAL   SH(10,10)

      DO I=1,KD
      DO K=1,KD
      SPUMA(I,K)=0.
      END DO
      END DO

      DO I=1,KD
      DO J=1,KD
      DO K=1,KD
      SPUMA(I,J)=SPUMA(I,J)+DBLE(SH(I,K))*STOIN(K,J)
      END DO
      END DO
      END DO

      PV=0.

      DO I=1,KD
      PV=PV+SNGL(SPUMA(I,I))
      END DO

      R E T U R N

      END
```

ANHANG B5

Verzeichnis der vorhandenen Permutationstest-Programme

PITU	unverbundenes 2-Stichprobenproblem
PITV	verbundenes 2-Stichprobenproblem
CR1F	unverbundenes k-Stichprobenproblem
RB1F	verbundenes k-Stichprobenproblem
SP	Split-Plot Versuchsplan
RB2F	2-faktorieller vollständiger Blockplan
SSP	Split-Split-Plot Versuchsplan
SPLITPLOTQ1Q2	Quadratische Formen als Teststatistiken im Split-Plot Versuchsplan
RB2FQ	Quadratische Formen als Teststatistiken im zweifaktoriellen Blockplan
SSP_Q1Q2Q3	Quadratische Formen als Teststatistiken im Split-Split-Plot Versuchsplan
SP_PROFIL	Test auf Profilparallelität (Split-Plot Plan als CR1F-MANOVA Problem)
SSP_PROFIL	Test auf verallgemeinerte Profilparallelität (Split-Split-Plot Plan als CR1F-MANOVA Problem)
STROE1	Shift-Algorithmus nach Streitberg & Röhmel (1986) für das verbundene 2-Stichprobenproblem
STROE2	2-dimensionaler Shift-Algorithmus nach Streitberg & Röhmel für das unverbundene 2-Stichprobenproblem
PT2	Algorithmus nach Pagano & Tritchler (1983) für 2-Stichprobenprobleme
JOHN2	p-Wert Approximation und Konfidenzintervall für den Effektparamter im unverbundenen 2-Stichprobenproblem nach John & Robinson (1983a)
CR1FPOWER	Exakte und naive Teststärkeschätzung im CR-k
RB1FPOWER	Exakte und naive Teststärkeschätzung im RB-k
GABRIEL_CR	Sequentielle multiple Testprozedur nach Petrondas & Gabriel(1983) für den CR-k
GABRIEL_RB	Sequentielle multiple Testprozedur nach Petrondas & Gabriel (1983) für den RB-k

WELSCH_CR	Sequentielle multiple Testprozedur nach Welsch (1977) für den CR-k
WELSCH_RB	Sequentielle multiple Testprozedur nach Welsch (1977) für den RB-k
MANY_ONE_CR	Multiple Testprozedur für die Vergleiche gegen eine Kontrolle im CR-k
MANY_ONE_RB	Multiple Testprozedur für die Vergleiche gegen eine Kontrolle im RB-k
FOUTZ_CR	Multiple Testprozedur nach Foutz (1981) für den CR-k
FOUTZ_RB	Multiple Testprozedur nach Foutz (1981) für den RB-k
FOUTZ_MANY_ONE_CR	Wie FOUTZ_CR für Vergleiche gegen eine Kontrolle
FOUTZ_MANY_ONE_RB	Wie Foutz_RB für Vergleiche gegen eine Kontrolle

PITUMANOVA	
PITVMANOVA	
CR1FMANOVA	
RB1FMANOVA	Multivariate Versionen der univariaten Tests
SPMANOVA	von oben
RB2FMANOVA	
SSPMANOVA	

SPMANOVA_PROFIL Test auf multivariate Profilparallelität

PITU_MULTI	
PITV_MULTI	
CR1F_MULTI	Multiple sequentielle Testprozeduren für die
RB1F_MULTI	abhängigen Variablen in einem der multivariaten
SP_MULTI	Versuchspläne von oben
RB2F_MULTI	

LEBENSLAUF

Name:	Klaus Willmes	
geboren:	11. September 1950 in Neheim-Hüsten, Kreis Arnsberg, als einziges Kind des Prokuristen Fritz Willmes und seiner Ehefrau Christa, geborene Städing	
Staats-angehörigkeit:	deutsch	
Familienstand:	ledig	
Schulbildung:	1961 - 1969	Franz-Stock-Gymnasium in Neheim-Hüsten (Abitur Juni 1969)
Studium:	1969 - 1974	Diplomstudiengang Mathematik RWTH Aachen (Abschluß: Diplom-Mathematiker, September 1974)
	1973 - 1979	Diplomstudiengang Psychologie RWTH Aachen (Abschluß: Diplom-Psychologe, August 1979)
bisherige Tätigkeiten:	05.73 - 09.74	Studentische Hilfskraft am Institut für medizinische Statistik und Dokumentation der RWTH Aachen (Prof. Dr. R. Repges)
	10.74 - 02.79	Wissenschaftliche Hilfskraft am selben Institut
	seit 03.79	Wissenschaftlicher Angestellter als Mitglied der interdisziplinären Arbeitsgruppe Neurolinguistik an der Abteilung Neurologie (Prof. Dr. K. Poeck), Mitarbeiter an den DFG-Forschungsprojekten Aphasie-Katamnese, Sprachliches Lernen bei Aphasie
	seit 04.83	Projektleiter in der DFG-Forschergruppe "Aphasie und kognitive Störungen"
	seit 1982	Mitglied im Advisory Board des Journal of Neurology
	seit 1986	Mitglied im Editorial Board der Zeitschrift Aphasiology
	seit 1987	Mitglied im Board of Associate Editors der Zeitschrift Neuropsychologia

offsetdruck
tel. (0241) 550467